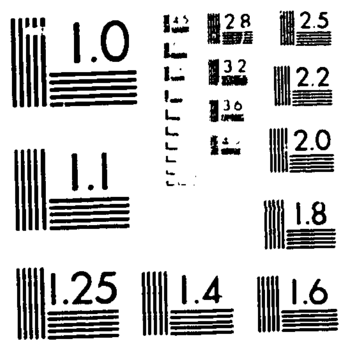
1.0
2.8
2.5
3.2
2.2
3.6
2.0
1.1
1.8
1.25
1.4
1.6

DOCUMENT RESUME

ED 174 633 TM 008 948

AUTHOR	Gustafsson, Jan-Eric; Lindstrom, Berner
TITLE	Analyzing ATI Data By Structural Analysis of Covariance Matrices.
INSTITUTION	Goteborg Univ., Molndal (Sweden). Inst. of Education.
SPONS AGENCY	National Swedish Board of Education, Stockholm.; Swedish Council for Research in the Humanities and Social Sciences, Stockholm.
PUB DATE	Apr 79
NOTE	64p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)
AVAILABLE FROM	Institute of Education, University of Goteborg, Fack, S-431 20 Molndal, Sweden
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Analysis of Covariance; *Aptitude; Classroom Research; Computer Programs; *Mathematical Models; *Multiple Regression Analysis; *Research Design
IDENTIFIERS	*Aptitude Treatment Interaction; *LISREL Computer Program

ABSTRACT

Joreskog has developed a very general and powerful model and computer program for analyzing relationships among variables, called LISREL. Structural analysis of covariance matrices by Joreskog's LISREL method is proposed as an alternative to regression methodology in the analysis of aptitude-treatment interaction (ATI) data. LISREL resolves some of the basic problems in regression analysis--such as biased estimates caused by unreliability of measures, and interpretative difficulties when there are multiple aptitudes and outcomes. LISREL models range in complexity: from regression models to models involving complex structural relations between latent aptitude and latent outcome variables. (Consideration is also given to the formulation of LISREL models when there are observational units at different levels of aggregation). (Author/JKS)

ANALYZING ATI DATA BY STRUCTURAL ANALYSIS OF COVARIANCE MATRICES [1)]

Jan-Eric Gustafsson
Berner Lindström

University of Göteborg

Paper presented at the Annual Meeting of the American Educational Research Association, San Fransisco, April 8-12, 1979. Requests for copies should be sent to Jan-Eric Gustafsson or Berner Lindström, Institute of Education, University of Göteborg, Fack, S-431 20 MÖLNDAL, Sweden.

1) The research presented in this paper has been supported financially by the Swedish Council for Research in the Humanities and Social Sciences and by the National Board of Education in Sweden. We also wish to express our gratitude to Dag Sörbom, who has contributed invaluable advice during our work with the paper.

ABSTRACT

Structural analysis of covariance matrices by Jöreskog's LISREL method is proposed as an alternative to regression methodology in the analysis of ATI data. It is shown how LISREL resolves some of the basic problems in regression analysis, such as biased estimates caused by unreliability of measures and interpretative difficulties when there are multiple aptitudes and outcomes. The formulation of LISREL models ranging in complexity is examplified: from regression models to models involving complex structural relations between latent aptitude and latent outcome variables. Consideration is also given to the formulation of LISREL models when there are observational units at different levels of aggregation.

Introduction

Cronbach (1957) argued that the experimental and correlational lines of behavioral science should be merged into one, taking jointly into account differences between treatment conditions and individual differences. Within educational psychology such research has been carried out in investigations of aptitude-treatment interactions (ATIs,Cronbach & Snow, 1977).

Not unlike many other lines of research within educational psychology, the results in ATI research tend to be inconsistent. Cronbach (1975) explained the inconsistencies as being caused by higher-order interactions, just as the presence of ATIs can be invoked to explain the inconsistent results from the experimental research on teaching methods (cf. Gustafsson, 1975).

Undoubtedly higher-order interactions account for many seemingly contradictory ATI-results. But other explanations can be invoked as well, and in this paper we will point out one: the methods of statistical analysis that have been used.

Multiple regression analysis, in one form or another, has been the standard method of analysis of ATI data, with heterogeneous within-treatment regressions of outcome on aptitude signifying ATI. But, as will be elaborated upon below, errors of measurement in the aptitude variables enter bias into the estimates of regression coefficients and thereby also into the description and testing of ATI effects (cf. Cronbach & Snow, 1977, pp. 33-34).

Another complication arises in regression analysis when there are several outcome and/or aptitude variables. Regression analysis is univariate in the sense that only one outcome variable is treated at a time, so they must be analyzed separately. Several aptitude variables can be analyzed at the same time in multiple regression analysis but for each aptitude variable a separate regression coefficient is estimated. Thus a great many regression coefficients are estimated and tested, which makes for chance significancies and tends to give rise to complex patterns of results which are hard to interpret.

These two problems, unreliability of the aptitude variables and large sets of variables, also tend to appear together and magnify each other, since short tests are often used to cover a wider range of aptitude variables.

The main purpose of the paper is, however, to point at the availability of an alternative method of analysis with which these problems can often be solved -- the LISREL method of Jöreskog and Sörbom.

The plan of the paper is as follows: Following a brief description of the LISREL approach, the problem of errors of measurement in the aptitude variables is discussed, along with different methods for solving it. Then examples of LISREL models of increasing complexity are formulated in relation to empirical ATI-studies, to illustrate how the method can be used. Finally advantages and disadvantages of this alternativ method of analysis are discussed.

1. LISREL

The abbreviation LISREL stands for linear structural relations and it is a model of high generality in which many other statistical models can be found as special cases (cf. Jöreskog & Sörbom, 1978, pp. 2-3) LISREL was introduced by Jöreskog (1973, 1977), and a description of the model, and a computer program of the same name (LISREL IV) is given by Jöreskog and Sörbom (1978, cf. also Jöreskog & Sörbom, 1976, 1977). LISREL includes as special cases the methods for analysis of covariance structures developed by Jöreskog (1969, 1970, 1971, 1974). Here only a very sketchy description of LISREL can be given, and for a full account the reader should consult the references.

The LISREL model consists of two parts: the measurement models for the dependent and independent variables, in which latent variables (common factors) are defined in terms of observed variables, and the linear structural equation model, in which the relations between the latent variables are specified.

The measurement models are factor analysis models in which a smaller set of latent variables (factors) are supposed to account

for the relations between the observed variables, and which are used to describe the measurement characteristics of the observed variables. There are two sets of observed variables $\underset{\sim}{y}'=(y_1,y_2,\ldots, y_p)$ and $\underset{\sim}{x}'=(x_1,x_2,\ldots,x_q)$, corresponding to dependent (outcome) and independent (aptitude) variables respectively and two sets of latent variables $\underset{\sim}{\eta}'=(\eta_1,\eta_2,\ldots,\eta_m)$ and $\underset{\sim}{\xi}'=(\xi_1,\xi_2,\ldots,\xi_n)$, corresponding to dependent and independent latent variables, respectively. There also are vectors specifying the unique parts (errors of measurement and specificity) of the $\underset{\sim}{y}$ and $\underset{\sim}{x}$ variables, $\underset{\sim}{\varepsilon}'= (\varepsilon_1, \varepsilon_2,\ldots,\varepsilon_p)$ and $\underset{\sim}{\delta}'= (\delta_1,\delta_2,\ldots,\delta_q)$.

The relations between the latent and the observed independent variables are specified in $\underset{\sim}{\Lambda}_x$, which is a factor loading matrix of order q x n for the regression of the $\underset{\sim}{x}$ variables on the $\underset{\sim}{\xi}$ variables, and the relations between the latent and the observed dependent variables are specified in the corresponding factor loading matrix $\underset{\sim}{\Lambda}_y$, of order p x m.

The measurement model for the $\underset{\sim}{x}$ variables is written:

(1) $$\underset{\sim}{x} = \underset{\sim}{\Lambda}_x\underset{\sim}{\xi} + \underset{\sim}{\delta},$$

and for the $\underset{\sim}{y}$ variables it is written:

(2) $$\underset{\sim}{y} = \underset{\sim}{\Lambda}_y\underset{\sim}{\eta} + \underset{\sim}{\varepsilon}.$$

The structural equation model specifies the causal relationships among the latent variables and to represent these,two parameter matrices are used: $\underset{\sim}{\Gamma}$ which is a coefficient matrix of order m x n for the structural relations between the $\underset{\sim}{\xi}$ and the $\underset{\sim}{\eta}$ variables; and $\underset{\sim}{\beta}$ which is a coefficient matrix of order m x m for the structural relations among the $\underset{\sim}{\eta}$ variables. The residuals (disturbance terms or errors in equations) in the dependent variables are represented with the vector: $\underset{\sim}{\zeta}'=(\zeta_1,\zeta_2,\ldots\zeta_m)$.

The system of linear structural relations has the form:

(3) $$\underset{\sim}{\beta}\underset{\sim}{\eta} = \underset{\sim}{\Gamma}\underset{\sim}{\xi} + \underset{\sim}{\zeta}.$$

The following covariance matrices must also be defined:

Θ_δ is a diagonal or symmetric matrix of order q x q containing the covariance matrix for the unique parts of the x variables.

Θ_ε is a diagonal or symmetric matrix of order p x p containing the covariance matrix for the unique parts of the y variables.

Φ is a diagonal or symmetric matrix of order n x n containing the covariance matrix of the ξ variables.

Ψ is a diagonal or symmetric matrix of order m x m for the covariance of the residuals.

Thus, in LISREL it is not necessarily assumed that the errors of measurement in the independent and dependent variables are uncorrelated with each other. It should also be pointed out that it is possible to specify LISREL models which allow estimation of covariances between errors of measurement in the independent and dependent variables; this can be effected through specifying a model in y variables only (cf. Example 7, below). It is assumed, however, that the errors of measurement are uncorrelated with ξ, η and ζ.

It can be shown (cf. Jöreskog & Sörbom, 1978, p. 5) that if a set of observational data can be described with the equations (1), (2) and (3), and if the other assumptions are fulfilled, the covariance matrix Σ of order (p+q) x (p+q) of the observed dependent and independent variables is:

$$(4) \qquad \Sigma = \begin{bmatrix} \Lambda_y(\beta^{-1}\Gamma\Phi\Gamma'\beta'^{-1}+\beta^{-1}\Psi\beta'^{-1})\Lambda_y'+\Theta_\varepsilon & \Lambda_y\beta^{-1}\Gamma\Phi\Lambda_x' \\ \Lambda_x\Phi\Gamma'\beta'^{-1}\Lambda_y' & \Lambda_x\Phi\Lambda_x'+\Theta_\delta \end{bmatrix}$$

In specifying a LISREL model it is necessary to specify the nature of each element in the matrices $\Lambda_x, \Lambda_y, \Gamma, \beta, \Phi, \Psi, \Theta_\delta$ and Θ_ε (the elements will be referred to with small Greek letters). The elements can be of three different kinds: a fixed parameter, i.e. the parameter is assigned a given value; a free parameter, i.e. the parameter is to be estimated; and a constrained parameter, i.e. the parameter is to be estimated but it is constrained to be equal to one or more other parameters.

From the relations (1)-(3) it would seem that LISREL is subject to a major limitation -- the means of the latent or the observed variables are not included in the model. In terms of regression analysis that would correspond to regression models without the intercept parameter, and to make a complete evaluation of ATI effects it is necessary to include the intercept as well. Sörbom (1974, 1976, 1978) has formulated models which do allow hypotheses on the means, and which come very close to the LISREL model. As has been shown by Sörbom (1979) these models can in fact be estimated with the LISREL program, using a special specification, so in reality LISREL does allow estimation and testing of the intercept parameter. In most of the examples to be presented below the intercept will however not be included, but the procedure is illustrated in Example 7 below.

So far LISREL has been presented as if there was one group (population) of persons only, but in ATI applications there always are two or more groups of persons, each having had a different treatment. LISREL handles any number of groups, however, and the presentation given above is easily generalized, through adding a superscript (i, i=1,...,g) indicating to which of g groups a parameter or a matrix of parameters refers. Thus, for example, the matrix of coefficients of structural relations between independent and dependent latent variables in the ith group is referred to as $\underset{\sim}{\Gamma}^{(i)}$. If a parameter or a matrix of parameters is constrained to be equal in all groups an asterisk (*) is used to denote that, i.e. $\underset{\sim}{\Gamma}^{(*)}$.

The values of the non-fixed parameters in the LISREL model can be estimated from the sample covariance matrices. However, to obtain any estimates it is necessary that the model is identified. The problem of identifiability can be defined in the following way:

> Identifiability depends on the choice of the model and on the specification of fixed, constrained, and free parameters. Under a given specification, a given structure $\underset{\sim}{\Lambda}_y, \underset{\sim}{\Lambda}_x, \underset{\sim}{B}, \underset{\sim}{\Gamma}, \underset{\sim}{\Phi}, \underset{\sim}{\Psi}, \underset{\sim}{\Theta}_\varepsilon, \underset{\sim}{\Theta}_\delta$ generates one and only one $\underset{\sim}{\Sigma}$ but there may be several structures generating the same $\underset{\sim}{\Sigma}$. If two or more structures generate the same $\underset{\sim}{\Sigma}$, the structures are said to be equivalent. If a parameter has the same value in all equivalent structures, the parameter is said to be identified. If all parameters of the model are identified, the whole model is

said to be identified. (Jöreskog & Sörbom, 1978, p.9).

For some special cases there are general rules for determining whether a specific model is identified or not (e.g. Werts, Jöreskog & Linn, 1973; Wiley, 1973) but in most instances that is not the case. The LISREL IV program has, however, the capability of detecting if a model is not identified (cf. Jöreskog & Sörbom, 1978, pp. 10-11).

In an identified model the values of the non-fixed parameters can be estimated with maximum likelihood methods. It is assumed that the distribution of the observed variables is sufficiently well described by the moments of the first and second orders, which in particular holds true when the observed variables have a multinormal distribution.

Each analysis of a fully identified model not only yields estimates of parameters but also an overall chi-square test of the goodness of fit of the model, along with standard errors of the estimated parameters. As a help in modification of a poorly fitting model, the first derivatives with respect to the fixed parameters are also computed (cf. Sörbom, 1975).

Through computing the differences between the values of the test statistics obtained with more and less constrained models, i.e. models differing as to the number of parameters estimated, it is also possible to test the significance of subsets of parameters. Consider the following concrete example: A model is estimated for two groups in which $\underset{\sim}{\Gamma}^{(1)}$ and $\underset{\sim}{\Gamma}^{(2)}$ are not constrained to be equal. The test of fit gives χ^2_1 with df_1 degrees of freedom. Then a model is specified in which $\underset{\sim}{\Gamma}^{(*)}$ is estimated instead, which will have χ^2_2 with df_2 degrees of freedom. The test statistic $\chi^2_2 - \chi^2_1$ then is chi-square distributed with $df_2 - df_1$ degrees of freedom, and the test is, of course, a test of the equality of the coefficients of structural relations within treatments. This test is the one which above all is of interest in ATI applications. In the same way other parameter matrices, or subsets of parameter matrices, can be tested.

Within the LISREL framework it is possible to formulate a wide range of different models, which can accomodate most of the designs used in ATI research. Several such models will be presented below, and we will start with a model which is identical to regression analysis.

2. Univariate regression on observed aptitude variables

Multiple regression analysis (MR) can be handled by LISREL as a special case. In MR the $\underset{\sim}{x}$ variables are considered fixed, i.e. $\underset{\sim}{\Lambda}_x = \underset{\sim}{I}$ and $\underset{\sim}{\Theta}_\delta = \underset{\sim}{0}$, so the measurement model for the independent variables reduces to $\underset{\sim}{x} = \underset{\sim}{\xi}$. In MR there is only one dependent variable, y, and it is similarly assumed that y=η. The structural relations model for treatment group i then reduces to:

$$y^{(i)} = \underset{\sim}{\Gamma}^{(i)} \underset{\sim}{x}^{(i)} + \zeta^{(i)}, \tag{5}$$

i.e. the ordinary MR model without the intercept parameter. The intercept can be estimated, however, through adding another x variable, which for all persons has the value 1, and through analyzing instead of the covariance matrix the matrix of moments about zero (cf. Jöreskog & Sörbom, 1978, pp. 8-9).

The $\underset{\sim}{\Gamma}^{(i)}$ parameters of the LISREL model (5) can be estimated in the usual way, and they can be tested for equality in the treatment groups. This test gives the same result as the ordinary F-test of homogeneity of regression. It is, however, not possible to compute a meaningful goodness-of-fit test within each group since a regression model of this kind always fits the data perfectly.

When there are errors of measurement in the $\underset{\sim}{x}$ variables the LISREL model above is misspecified. The attenuating effects of errors of measurement on the estimates of regression coefficients are well known (e.g. Härnqvist, 1968; Cronbach, Gleser, Nanda & Rajaratnam, 1972; Cronbach & Snow, 1977, pp. 33-35) but we will nevertheless treat that in detail for some simple cases, to illustrate the analytical power of LISREL and

also to make clear the quite drastic effects which errors of measurement may have on the pattern of results from ATI studies.

We first consider the case when there is 1 y variable and 1 x variable only. The $\underset{\sim}{\Gamma}^{(i)}$ matrix then contains one element only, $\gamma^{(i)}$. Suppose that the x variable has the measurement model:

(6) $$x^{(i)} = \xi^{(i)} + \delta^{(i)}$$

Since it is assumed that the errors of measurement are uncorrelated with ξ it follows that the observed variance $\sigma_{xx}^{(i)} = \sigma_{\xi\xi}^{(i)} + \Theta_{\delta}^{(i)}$. Using (4) it is easily shown that $\sigma_{xy}^{(i)} = \gamma^{(i)} \sigma_{\xi\xi}^{(i)}$.

If $\Theta_{\delta}^{(i)} = 0$ we can estimate $\gamma^{(i)}$ through taking $\frac{\sigma_{xy}^{(i)}}{\sigma_{xx}^{(i)}}$ which is also the estimator used in regression analysis. It is obvious, however, that if $\Theta_{\delta}^{(i)} \neq 0$ a biased estimate will be obtained if this estimator is used. If $\Theta_{\delta}^{(i)}$ is known, however, an unbiased estimate is obtained if the estimator $\gamma^{(i)} = \frac{\sigma_{xy}^{(i)}}{\sigma_{xx}^{(i)} - \Theta_{\delta}^{(i)}}$ is used instead. Using $\rho_{x}^{(i)}$ to denote the ratio $\frac{\sigma_{\xi\xi}^{(i)}}{\sigma_{xx}^{(i)}}$, or the reliability, the biased estimate of $\gamma^{(i)}$ can also be made unbiased through using the correction $\frac{\gamma^{(i)}}{\rho_{x}^{(i)}}$.

When there are 2 treatment groups the difference $\Delta_{\gamma} = \gamma^{(1)} - \gamma^{(2)}$ defines the strength of the ATI effect. If however, there are errors of measurement in the aptitude variable and the reliability is the same in the 2 treatment groups, the difference observed if the biased estimator is used will in the long run be:

(7) $$\Delta_{\gamma} = \rho_{x}^{(*)} (\gamma^{(1)} - \gamma^{(2)})$$

Since ρ_x is never greater than unity, unreliability of an aptitude variable systematically lowers the possibility of detecting ATI effects.

As was shown by Cronbach and Snow (1977, pp. 33-34) the attenuation also affects the crossover point of the regressions when the interaction is disordinal, and it may even change the ordinality of the interaction. Cronbach and Snow (1977, cf. Cronbach, 1976) also pointed out that if the reliabilities differ between treatment groups, the attenuated coefficients may well be identical, while at the same time the unattenuated ones differ.

It can thus be concluded that in regression analysis of one aptitude variable only, errors of measurement in that variable tend to diminish the possibility of detecting ATI effects, with the effect of unreliability being a linear function of the amount of unreliability, at least as long as the reliability is the same in all groups. Errors of measurement in the outcome variable does not have any biasing effects on the estimates, however, and if there is such variance it can be absorbed by $\psi^{(i)}$.

In MR the biasing effects of errors of measurement in the aptitude variables generally tend to be both stronger and more complex, which is illustrated below for the special case of 2 aptitude variables. Suppose that the measurement model for the aptitude variables is:

$$(8) \qquad \begin{aligned} x_1^{(i)} &= \xi_1^{(i)} + \delta_1^{(i)} \\ x_2^{(i)} &= \xi_2^{(i)} + \delta_2^{(i)} \end{aligned}$$

The errors of measurement are here assumed to be uncorrelated with each other. In explicit notation the structural equation model is:

$$(9) \qquad y^{(i)} = \gamma_1^{(i)}\xi_1 + \gamma_2^{(i)}\xi_2 + \zeta^{(i)}$$

From (4), (8) and (9) it follows that the covariance matrix generated for the ith group is:

$$
(10)\qquad \underset{\sim}{\Sigma}^{(i)} = \begin{bmatrix}
\gamma_1^{(i)^2}\sigma_{\xi_1\xi_1}^{(i)} + 2\gamma_1^{(i)}\gamma_2^{(i)}\sigma_{\xi_1\xi_2}^{(i)} + \gamma_2^{(i)^2}\sigma_{\xi_2\xi_2} + \psi^{(i)} & & \\
\gamma_1^{(i)}\sigma_{\xi_1\xi_1}^{(i)} + \gamma_2^{(i)}\sigma_{\xi_1\xi_2} & \sigma_{\xi_1\xi_1}^{(i)} + \Theta_{\delta_1}^{(i)} & \\
\gamma_2^{(i)}\sigma_{\xi_2\xi_2}^{(i)} + \gamma_1^{(i)}\sigma_{\xi_1\xi_2}^{(i)} & \sigma_{\xi_1\xi_2}^{(i)} & \sigma_{\xi_2\xi_2}^{(i)} + \Theta_{\delta_2}^{(i)}
\end{bmatrix}
$$

The observed covariance matrix is:

$$
(11)\qquad \underset{\sim}{S}^{(i)} = \begin{bmatrix}
\sigma_{yy}^{(i)} & & \\
\sigma_{x_1y}^{(i)} & \sigma_{x_1x_1}^{(i)} & \\
\sigma_{x_2y}^{(i)} & \sigma_{x_1x_2}^{(i)} & \sigma_{x_2x_2}^{(i)}
\end{bmatrix}
$$

Assuming $\underset{\sim}{\Theta}_{\delta}^{(i)} = \underset{\sim}{0}$ the ordinary MR estimation equations are easily derived from (10) and (11):

$$
(12)\qquad \begin{cases}
\gamma_1^{(i)} = \dfrac{\sigma_{x_1y}^{(i)}\,\sigma_{x_2x_2}^{(i)} - \sigma_{x_1x_2}^{(i)}\,\sigma_{x_2y}^{(i)}}{\sigma_{x_1x_1}^{(i)}\,\sigma_{x_2x_2}^{(i)} - \sigma_{x_1x_2}^{(i)}\,\sigma_{x_1x_2}^{(i)}} \\[3ex]
\gamma_2^{(i)} = \dfrac{\sigma_{x_2y}^{(i)}\,\sigma_{x_1x_1}^{(i)} - \sigma_{x_1x_2}^{(i)}\,\sigma_{x_1y}^{(i)}}{\sigma_{x_1x_1}^{(i)}\,\sigma_{x_2x_2}^{(i)} - \sigma_{x_1x_2}^{(i)}\,\sigma_{x_1x_2}^{(i)}}
\end{cases}
$$

If the reliabilities, $\rho_1^{(i)}$ and $\rho_2^{(i)}$, are known, the bias introduced in (12) by $\underset{\sim}{\Theta}_{\delta}^{(i)} \neq \underset{\sim}{0}$ can be avoided through using instead the estimator:

$$
(13)\quad
\begin{cases}
\gamma_1^{(i)} = \dfrac{\sigma_{x_1y}^{(i)}\,\sigma_{x_2x_2}^{(i)}\,\rho_2^{(i)} - \sigma_{x_1x_2}^{(i)}\,\sigma_{x_2y}^{(i)}}{\sigma_{x_1x_1}^{(i)}\,\sigma_{x_2x_2}^{(i)}\,\rho_1^{(i)}\rho_2^{(i)} - \sigma_{x_1x_2}^{(i)}\,\sigma_{x_1x_2}^{(i)}} \\[2em]
\gamma_2^{(i)} = \dfrac{\sigma_{x_2y}^{(i)}\,\sigma_{x_1x_1}^{(i)}\,\rho_1^{(i)} - \sigma_{x_1x_2}^{(i)}\,\sigma_{x_1y}^{(i)}}{\sigma_{x_1x_1}^{(i)}\,\sigma_{x_2x_2}^{(i)}\,\rho_1^{(i)}\rho_2^{(i)} - \sigma_{x_1x_2}^{(i)}\,\sigma_{x_1x_2}^{(i)}}
\end{cases}
$$

From (13) follows that in MR the attenuation effects are non-linear functions of the amounts of error of measurement and the covariance between the aptitude variables.

Since the effects are quite complex any general statements cannot be made, but Figure 1 illustrates the effects in one particular hypothetical ATI application. In this figure the expected observed differences $\hat{\gamma}_1^{(1)} - \hat{\gamma}_1^{(2)}$ and $\hat{\gamma}_2^{(1)} - \hat{\gamma}_2^{(2)}$ have been plotted as a function of $\rho_1^{(*)}$, with $\rho_2^{(*)}$ taken to be .90, $\gamma_1^{(1)} = .20$, $\gamma_2^{(1)} = .40$, $\gamma_1^{(2)} = .40$, $\gamma_2^{(2)} = .20$ and $\sigma_{\xi_1\xi_2} = .50$. Even when both aptitude variables have reliabilities as high as .90 the observed difference between the coefficients is only about 80 percent of the true difference. If x_2 has a high reliability of .90 and x_1 has a lower reliability of .50 the observed difference between the attenuated coefficients for x_2 is, in this particular case, 70 percent of the true difference, while the observed difference for x_1 is only 35 percent of the true difference. Had in the example the true correlation between the aptitude variables been higher, even more marked effects would have been found.

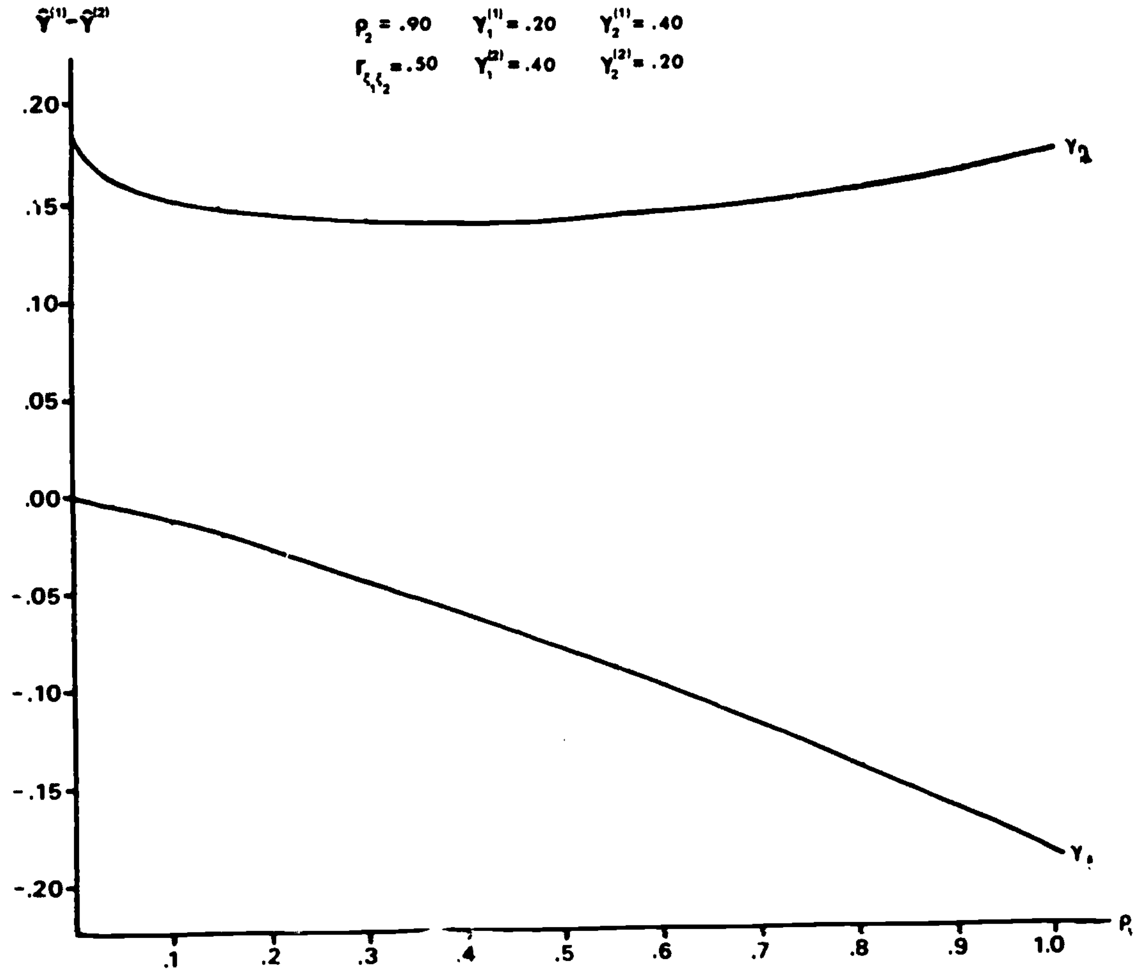

Figure 1 The differences between the expected attenuated within-treatment regression coefficients as a function of the reliability of one of the aptitude variables.

This is only one example, and other examples with even more drastic effects could easily be constructed. It may even happen that the observed difference between the within-treatment coefficients has the wrong sign.

Obviously unreliability in the aptitude variables has so marked an effect on the results from ATI studies that unattenuated coefficients should never be interpreted unless the reliability of all the aptitude variables in well above .90 and their correlation is low. Cronbach and Snow (1977, pp. 33-37, p.515) urged ATI researchers to disattenuate the regression coefficients. Only few researchers have followed their advice, however, so the bias entered by errors of measurement in the aptitude variables is likely to account for a great many seemingly contradictory ATI results.

Cronbach et al. (1972, pp. 287-302) describe methods useful in corrections for attenuation, which methods do not necessarily assume that the errors of measurement are uncorrelated. We shall in relation to a concrete example show how correction for attenuation can also be effected with LISREL.

EXAMPLE 1: Disattenuating regressions

Most of the examples to be presented will employ real data, but here we will use generated data. For each of 2 treatment groups a LISREL model was constructed and the models are shown in Figure 2.

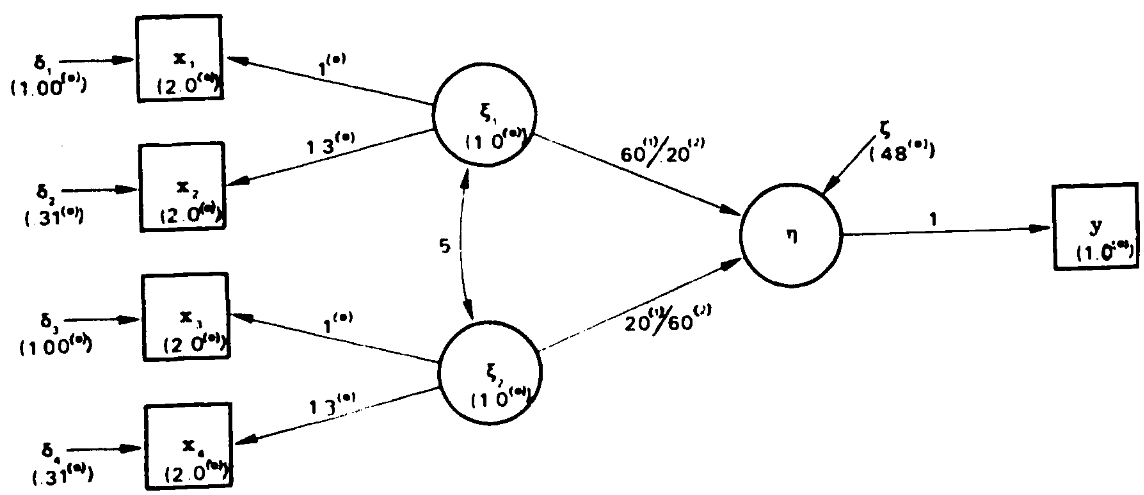

Figure 2 The model for the generated data

In the figure the following symbols have been used (cf. Jöreskog & Sörbom, 1978): Latent variables are enclosed in circles, observed variables are enclosed in squares, and errors of measurement and disturbance terms are included without being enclosed. A straight one-way arrow indicates a causal influence of one variable on another and curved two-way arrows indicate covariance between variables without any causal interpretation.

In the models variances are given in pharanthesis. Thus, for both groups the measurement model for the aptitude variables is identical, there being 2 ξ-variables, each being measured by 2 observed x-variables, with reliabilities .50 and .845. In group

1 ξ_1 is assumed to have a strong relation to an observed outcome variable, while in group 2, ξ_2 is assumed to be strongly related to outcome.

From Figure 2 follows directly that the following parameter specification was used:

$$\Lambda_x^{(*)} = \begin{pmatrix} 1.0 & 0.0 \\ 1.3 & 0.0 \\ 0.0 & 1.0 \\ 0.0 & 1.3 \end{pmatrix}, \quad \Phi^{(*)} = \begin{pmatrix} 1.0 & \\ .5 & 1.0 \end{pmatrix}$$

$$\Theta_\delta^{(*)} = \text{diag} \; (1.0, .31, 1.0, .31), \; \Gamma^{(1)} = (.60, .20),$$

$$\Gamma^{(2)} = (.20, .60) \; \Psi^{(*)} = (.48)$$

Inserting the parameter matrices in (4) the following covariance matrices are obtained:

$$\Sigma^{(1)} = \begin{pmatrix} 1.00 & & & & \\ .70 & 2.00 & & & \\ .91 & 1.30 & 2.00 & & \\ .50 & .50 & .65 & 2.00 & \\ .65 & .65 & .845 & 1.30 & 2.00 \end{pmatrix}$$

$$\Sigma^{(2)} = \begin{pmatrix} 1.00 & & & & \\ .50 & 2.00 & & & \\ .65 & 1.30 & 2.00 & & \\ .70 & .50 & .65 & 2.00 & \\ .91 & .65 & .845 & 1.30 & 2.00 \end{pmatrix}$$

Using the GGNRM routine in the IMSL libarary of computer subroutines, data with a multivariate normal distribution was generated according to these Σ-matrices. There were 150 subjects in each treatment. The following sample covariance matrices were obtained:

$$\underset{\sim}{S}^{(1)} = \begin{bmatrix} 1.12 & & & & \\ .84 & 2.03 & & & \\ 1.16 & 1.44 & 2.29 & & \\ .63 & .48 & .80 & 1.91 & \\ .75 & .56 & 1.03 & 1.42 & 2.13 \end{bmatrix}$$

$$\underset{\sim}{S}^{(2)} = \begin{bmatrix} 1.03 & & & & \\ .52 & 2.15 & & & \\ .53 & 1.47 & 2.14 & & \\ .83 & .68 & .66 & 2.07 & \\ .95 & .95 & .98 & 1.56 & 2.26 \end{bmatrix}$$

Having data on 4 x-variables it would of course be natural to analyze them all in one MR equation for each treatment. But with data such as these this is not possible since after correction for attenuation some of the variables are so highly intercorrelated that problems of multicollinearity would occur. Thus the variables will be treated 2 at a time.

We consider first estimation and testing of the attenuated coefficients. The regression model for x-variables 1 and 3 is shown in Figure 3. To obtain the estimates the following specification was used:

$$\underset{\sim}{\Lambda}_x^{(*)} = \underset{\sim}{I},\ \underset{\sim}{\Theta}_\delta^{(*)} = \underset{\sim}{0},\ \underset{\sim}{\Phi}^{(i)} = S_x^{(i)},\ \underset{\sim}{\Gamma}^{(i)} = \left[\gamma_1^{(i)}, \gamma_2^{(i)}\right],$$

$$\underset{\sim}{\Lambda}_y^{(*)} = [1],\ \Theta_\varepsilon^{(*)} = [0],\ \underset{\sim}{\beta}^{(*)} = [1],\ \underset{\sim}{\Psi}^{(i)} = \left[\Psi^{(i)}\right]$$

where $S_x^{(i)}$ denotes the sample covariance matrix for the x-variables in the ith group. To test the equality of the regresson coefficients in the treatment groups the $\Gamma^{(i)}$ matrices were in another model constrained to be equal.

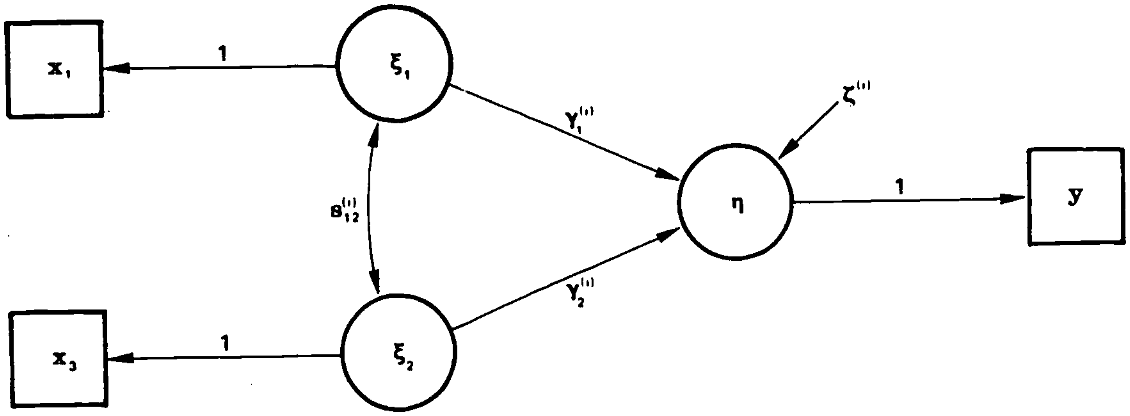

Figure 3 The model used to estimate the attenuated coefficients for the generated data.

To correct for attenuation we used the known population values of $\underset{\sim}{\Theta}_\delta^{(*)}$ and to obtained the unattenuated coeffici ts $\underset{\sim}{\Theta}_\delta$ can be entered as a matrix of fixed parameter values. The LISREL model used to obtain the unattenuated coefficients is shown in Figure 4.

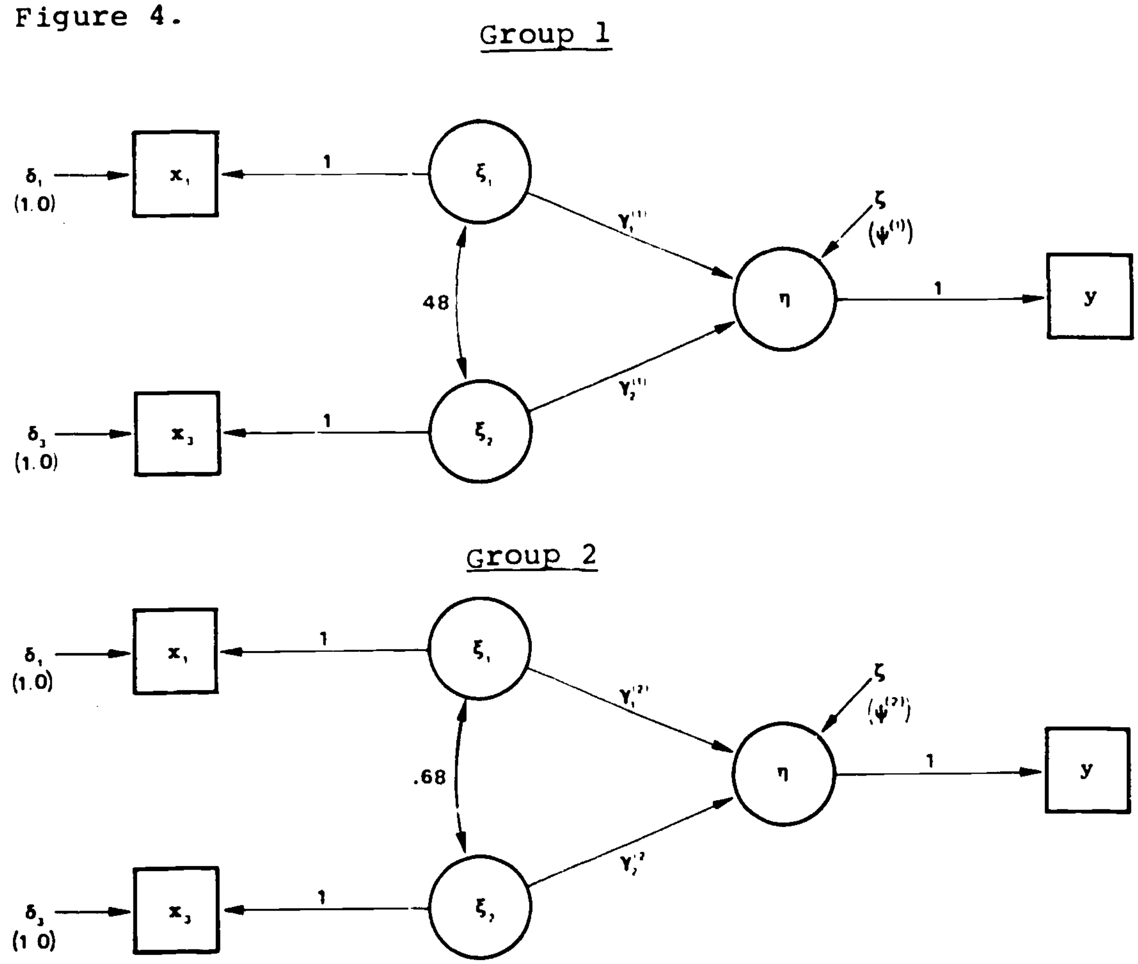

Figure 4 The model used to estimate the unattenuated coefficients for the generated data.

The following specification was used:

$$\Lambda_x^{(*)} = I, \quad \Theta_\delta^{(*)} = (1.0,\ 1.0), \quad \Phi^{(1)} = \begin{bmatrix} \phi_1^{(1)} & \\ .48 & \phi_2^{(1)} \end{bmatrix},$$

$$\Phi^{(2)} = \begin{bmatrix} \phi_1^{(2)} & \\ .68 & \phi_2^{(2)} \end{bmatrix}, \quad \Gamma^{(i)} = \begin{bmatrix} \gamma_1^{(i)}, \gamma_2^{(i)} \end{bmatrix}, \quad \Lambda_y^{(*)} = (1),$$

$$\Theta_\varepsilon^{(*)} = (0), \quad \beta^{(*)} = (1), \quad \Psi^{(i)} = \begin{bmatrix} \psi_1^{(i)} \end{bmatrix}$$

Thus, the diagonal of the $\Phi^{(i)}$ matrix contains free parameters and estimates of these parameters are of course estimates of the "true variances".

Estimates of the attenuated and disattenuated regression coefficients are presented in Table 1 for the 4 possible combinations of 2 x-variables at a time, 1 from each latent variable. The true coefficients are also presented, as are the results from the tests of equality of the estimated regression coefficients within the treatment groups.

Table 1

Estimates of regression coefficients for the generated data in example 1

	Attenuated coefficients					Unattenuated coefficients					True values			
x-variables	$\gamma_1^{(1)}$	$\gamma_1^{(2)}$	$\gamma_2^{(1)}$	$\gamma_2^{(2)}$	χ^2 1)	$\gamma_1^{(1)}$	$\gamma_1^{(2)}$	$\gamma_2^{(1)}$	$\gamma_2^{(2)}$	χ^2 1)	$\gamma_1^{(1)}$	$\gamma_1^{(2)}$	$\gamma_2^{(1)}$	$\gamma_2^{(2)}$
1 and 3	.36	.13	.24	.36	11.22	.65	-.00	.35	.77	11.63	.60	.20	.20	.60
1 and 4	.34	.07	.26	.39	15.56	.71	.09	.20	.44	12.44	.60	.20	.15	.46
2 and 3	.46	.14	.14	.36	26.76	.47	.02	.28	.76	17.65	.46	.15	.20	.60
2 and 4	.44	.07	.14	.39	32.32	.53	.04	.12	.47	31.92	.46	.15	.15	.46

1) The χ^2-value refers to the test of equality of the within-treatment coefficients, with 2 degrees of freedom.

The disattenuation of course has a considerable effect on the magnitude of difference between the within-treatment coefficients. The test of interaction, does, however, not result in any higher values on the test-statistic for the unattenuated coefficients than for the attenuated ones. This is due to the fact that the correction for attenuation increases the standard errors (cf. Bergman, 1971) and also to the fact that removal of the error variance from the $\underset{\sim}{\Phi}$ matrix causes a higher correlation between the $\underset{\sim}{\xi}$-variables, and the standard errors of the estimates of the $\underset{\sim}{\Gamma}$ coefficients are strongly affected by that.

The figures presented in Table 1 illustrate another point of some significance: the regression coefficients are not invariant for the x-variables measuring the same latent variable. This is because the true variance of the variables differ and coefficients of regression are not invariant over transformations of scale. Thus it is possible to find a significant interaction with one aptitude variable, while no interaction is found with another aptitude variable measuring the same latent variable but on another scale.

In this case we could use the population values as the fixed parameters of $\underset{\sim}{\Theta}_{\delta}^{(*)}$. In practice the population values are never known so some kind of estimate must be used. It must be pointed out, though, that it is not quite correct to estimate the reliabilities from the same sample as is used in the regression analysis. This is because the estimates are entered as fixed parameters in the LISREL model, but if they are estimated from the same sample they should in reality be treated as free parameters.

To conclude, we have seen how ordinary MR analysis can be handled by LISREL as well, and that the biasing effects of errors of measurement can be corrected for if there are estimates of the error variances (and possibly also error covariances). But we have also seen that apart from the problem of errors of measurement in the aptitude variables, MR meets with other problems: different aptitude variables which measure the same latent variable cannot be included in the same regression equation; the regression coefficients are affected by the scale on which the observed variables are measured; and for each observed variable one regression coefficient must be estimated which in tests of

interaction tends to inflate the chosen level of significance and tends to give rise to complex patterns of result, which are difficult to interpret. We have earlier concluded that the problems caused by use of attenuated coefficients are likely to account for many seemingly contradictory ATI-results -- the other problems listed above are likely to account for another share.

2. Univariate regression on latent aptitude variables

Even though it is possible to make corrections for attenuation to take into account the biasing effects of errors of measurement in the aptitude varibles, that solution is not the optimal one: the other problems associated with MR still remain and it is necessary to supply estimates of the error variances. But LISREL can, when there is more than 1 observed aptitude variable for each latent aptitude variable, be used to estimate the regression of observed or latent outcome variables on latent aptitude variables. Such analyses do not require that estimates of Θ_{δ} are supplied, instead estimates of Θ_{δ} are obtained.

EXAMPLE 2: A LISREL model for univariate regression on 2 latent aptitude variables

We will use the generated data in Example 1 to illustrate specification and estimation of parameters in a regression model with latent aptitude variables. The model is shown in Figure 5.

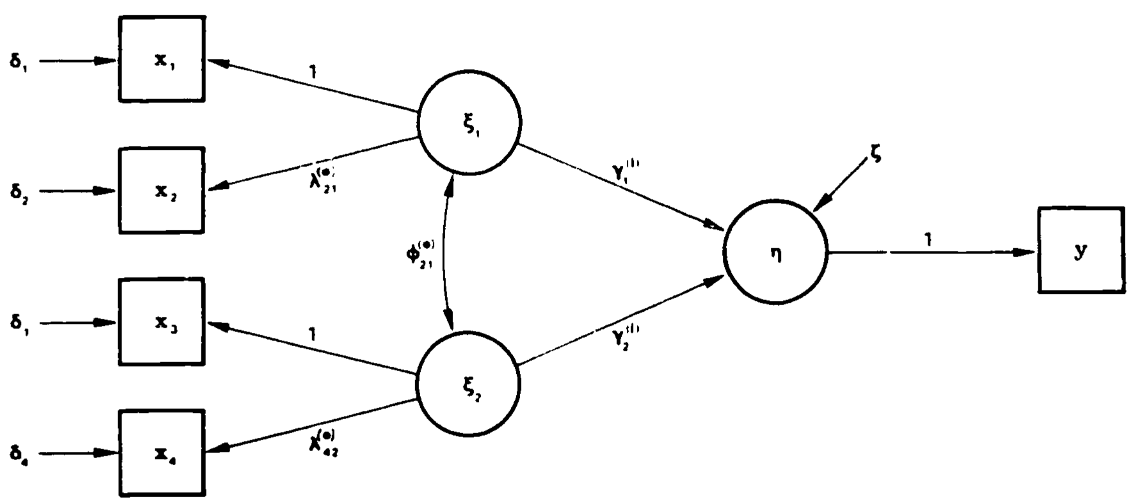

Figure 5 The model used to estimate the parameters for the generated data.

The following specification was used for i=1,2:

$$\underset{\sim}{\Lambda}_x^{(*)} = \begin{bmatrix} 1.0 & 0.0 \\ \lambda_{21}^{(*)} & 0.0 \\ 0.0 & 1.0 \\ 0.0 & \lambda_{42}^{(*)} \end{bmatrix} \quad \underset{\sim}{\Theta}_\delta^{(*)} = \left[\Theta_{\delta_{11}}^{(*)},\ \Theta_{\delta_{22}}^{(*)},\ \Theta_{\delta_{33}}^{(*)},\ \Theta_{\delta_{44}}^{(*)} \right],$$

$$\underset{\sim}{\Phi}^{(*)} = \begin{bmatrix} \phi_{11}^{(*)} & \\ \phi_{21}^{(*)} & \phi_{22}^{(*)} \end{bmatrix} \quad \underset{\sim}{\Gamma}^{(i)} = \left[\gamma_1^{(i)},\ \gamma_2^{(i)} \right],\ \underset{\sim}{\Lambda}_y^{(*)} = (1),$$

$$\underset{\sim}{\Theta}^{(*)} = (0)\ \underset{\sim}{\beta}^{(*)} = (1),\ \underset{\sim}{\Psi}^{(i)} = \left[\psi_1^{(i)} \right]$$

The specification follows quite straightforwardly from the specification used in generating the data. It should be pointed out, though, that for each ξ-variable, one of the elements of $\underset{\sim}{\Lambda}_x$ has been taken to be unity. This achieves a fixation of the scale of the latent variable to be equal to the scale of one of the observed variables. Any observed variable, and any constant can be used.

The parameter matrices in the measurement model for the aptitude variables i.e. $\underset{\sim}{\Lambda}_x^{(*)}$, $\underset{\sim}{\Phi}^{(*)}$ and $\underset{\sim}{\Theta}_\delta^{(*)}$ have been constrained to be equal in the groups, as they should be if the data have been successfully generated. If the parameters in the measurement model for the aptitude variables are constrained to be equal in the treatment groups, fewer parameters need to be estimated, which results in a more powerful analysis. Furthermore, if $\underset{\sim}{\Lambda}_x$ differs between the groups any interaction is of course difficult to interpret, as would also be the case if Φ differs between the groups. No problems of interpretation are caused by different $\underset{\sim}{\Theta}_\delta$, however, and if necessary one or more of the elements of $\underset{\sim}{\Theta}_\delta$ can be allowed to vary over the treatment groups. In the analysis of these data we also noted that the test of interaction was not invariant over transformations of the scales of the latent

variables when the $\underset{\sim}{\Lambda}_x$ matrix was not constrained to be equal in the treatment groups, which is another strong reason why in ATI research at least the $\underset{\sim}{\Lambda}_x$ matrix should be the same in all groups.

Table 2

Estimates of the free parameters for the generated data (Example 2)

	Group 1		Group 2
λ_{21}		1.29 (.11)	
λ_{42}		1.22 (.10)	
$\theta_{\delta_{11}}$		.95 (.11)	
$\theta_{\delta_{22}}$		.34 (.12)	
$\theta_{\delta_{33}}$		.77 (.10)	
$\theta_{\delta_{44}}$		.37 (.12)	
ϕ_{11}		1.14 (.17)	
ϕ_{21}		.62 (.10)	
ϕ_{22}		1.22 (.17)	
γ_1	.68 (.08)		.06 (.12)
γ_2	.18 (.07)		.60 (.09)
ψ_1	.41 (.06)		.54 (.09)

Standard errors are shown in parenthesis

The LISREL estimates of the free parameters are presented in Table 2, along with their standard errors. One of the estimates ($\Theta^{(*)}_{\delta_{33}}$) deviates more than 2 standard errors from the true value, which may be either an effect of chance or a reflection of an imperfection in the data generation routine.

A very good overall fit was found (χ^2=9.5, df=15, p<.85), and the fit is so good that there is no room for differences between

the treatments with respect to the parameters of the measurement model for the aptitude variables. Constraining the $\underset{\sim}{\Gamma}^{(i)}$ matrices to be equal, a very poor fit was found, however, (χ^2=43.0, df=17, <.00), and the test of overall interaction thus gives χ^2=33.5 which with 2 degrees of freedom is highly significant.

This approach to estimating and testing ATI effects effectively solves the problems associated with the MR approach: The regressions on latent variables can be estimated directly, the test of interaction is invariant under tranformations of the scales of the latent variables, and fewer coefficients need to be compared. Models of this kind can be used whenever it is reasonable to impose a factorial structure on the aptitude variables.

But even when there is 1 observed aptitude variable only for each hypothsized latent variable, it is sometimes possible to estimate regressions on the latent variables without using correction for attenuation. This can be done if "half-tests" are constructed by splitting the items in a measurement instrument into 2 roughly parallell halves which are both entered into the analysis. One drawback of this method is that it is not always possible to construct such half-tests for lack of item-level data, and another drawback is that power is lost. However, less power is lost if the half-tests are so well equated that the same λ parameters can be used for both. There is also another reason why as often as possible the same λ parameters should be used: if these are allowed to be different the estimates often tend to be unstable, with negative estimates of one or more of the elements of $\underset{\sim}{\Theta}_\delta$.

In most cases, however, development of the measurement model for the aptitude variables corresponds to imposing a factorial structure on the aptitude variables, and there are strong reasons to find one or more suitable measurement models in a first step, and not until later study relations between the latent aptitude variables and outcome variables.

Most of the aptitude variables used in ATI reserach have been used in other factor analytic investigations, so often previous research suggests which factorial structure to impose on the aptitude variables. Any such idea can be tested,thus using

LISREL to perform confirmatory factor analysis. Should a poor fit be found, the first derivatives with respect to the fixed parameters can be inspected to see how the model should be modified (Sörbom, 1975). Another possibility is, of course, to investigate the aptitude variables with exploratory factor analysis in a first step.

LISREL extends beyond ordinary factor analysis, however, in that the errors of measurement are not necessarily assumed to be uncorrelated. Such correlation can come about if the specific factors of 2 tests correlate, for example, or if the tests have been administered at a common occasion (cf. Cronbach et al., 1972). If there are such correlated errors of measurement, that would in ordinary factor analysis call for additional factors (cf. Sörbom, 1975).

As soon as there is a high correlation between the latent variables, the standard errors of the $\underset{\sim}{\Gamma}^{(i)}$ coefficients are high. Since the power of the test of ATI effects generally tends to be too small, it is important, if statistically significant interactions are sought, to keep the correlations between the latent variables low. One way of doing this is to let the errors of measurement for certain variables to correlated instead of adding further factors.

Example 3: Developing the measurement model for the aptitude variables

We will illustrate development of measurement models for the aptitude variables with some real data. A complete account of the ATI study from which the data are taken is given by Gustafsson (1979a) It included the following aptitude variables:

Opposites, a vocabulary test of verbal ability.
Figure series, a test of inductive or non-reasoning ability.
Metal folding and Cubes, which tests have been thought to measure spatial visualization (Vz) ability.
Figure and Flags, which tests are supposed to measure spatial orientation (SR-O) ability.
Since there was 1 verbal test only, the 40 items in Opposites were split in one half consisting of items with odd numbers and

one half consisting of items with even numbers.

The study included 2 treatment groups, called Reading and Listening with 159 and 155 subjects, respectively (see also Example 4, below). The covariance matrices for the aptitude variables are presented in Table 3.

Table 3

The covariance matrices for the aptitude variables in the Gustafsson (1979) study

	Reading						
	1	2	3	4	5	6	7
1. Opposites I	10.515						
2. Opposites II	8.331	11.981					
3. Metal folding	9.899	10.240	49.190				
4. Cubes	7.656	7.744	33.771	56.884			
5. Flags	30.143	30.383	117.735	109.607	862.442		
6. Figures	22.622	22.465	78.929	79.421	453.959	405.613	
7. Figure series	8.973	10.053	25.142	23.476	93.570	58.544	35.403

	Listening						
	1	2	3	4	5	6	7
1. Opposites I	8.525						
2. Opposites II	6.747	10.642					
3. Metal folding	5.539	5.379	36.649				
4. Cubes	3.712	5.439	16.718	32.913			
5. Flags	10.220	11.384	65.457	63.715	719.271		
6. Figures	9.163	11.368	40.167	48.082	315.977	302.637	
7. Figure series	5.671	8.131	14.284	13.048	46.094	31.241	35.891

In the Table the tests have been assigned numbers, and these numbers will be used to refer to the variables in the specification of LISREL models.

For reasons made explicit by Gustafsson (1976, 1979a) it was thought that performance on the Vz tests is to a large extent influenced by non-verbal reasoning ability, while the SR-0 tests were thought to be more "clean" indicators of spatial ability. This suggests testing of the 3-factor model shown in Figure 6.

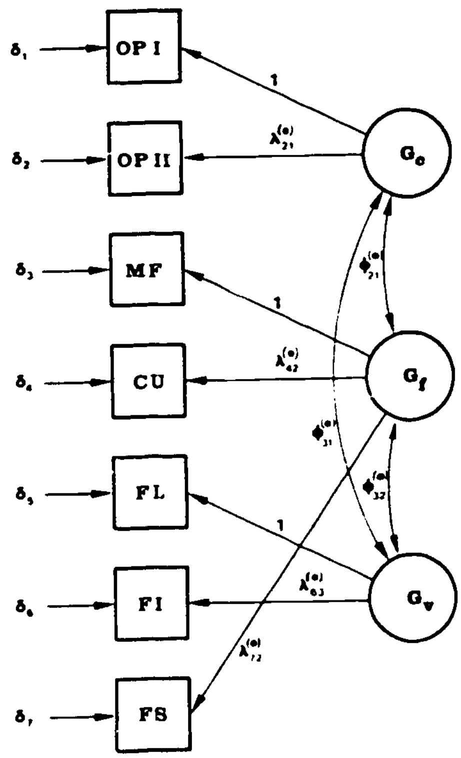

Figure 6 The original 3-factor model for the aptitude variables in the Gustafsson (1979a) study.

The specification of this model is as follows:

$$\underset{\sim}{\Lambda}_x^{(*)} = \begin{bmatrix} 1 & 0 & 0 \\ \lambda_{21}^{(*)} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_{42}^{(*)} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_{63}^{(*)} \\ 0 & \lambda_{72}^{(*)} & 0 \end{bmatrix}, \quad \underset{\sim}{\Phi}^{(*)} = \begin{bmatrix} \phi_{11}^{(*)} & & \\ \phi_{21}^{(*)} & \phi_{22}^{(*)} & \\ \phi_{31}^{(*)} & \phi_{32}^{(*)} & \phi_{33}^{(*)} \end{bmatrix},$$

$$\underset{\sim}{\Theta}_\delta^{(*)} = \text{diag}\left[\theta_{\delta_{11}}^{(*)}, \ldots, \theta_{\delta_{77}}^{(*)}\right]$$

The test of fit of this model gave a borderline significance ($\chi^2=53.7$, df=39, $p<.06$). Relaxing the constraints of equality of all the parameter matrices in the groups a somewhat better fit was found ($\chi^2=29.0$, df=22, $p<.14$), even though the test of the group difference is not significant.

Since none of the goodness-of-fit tests is significant it could be concluded that this simple 3-factor model is sufficient to account for the structure of the aptitude variables. But the p-values are quite low, and the samples are not large, so there is room for improvement of fit.

An analysis of the first-order derivatives with respect to the fixed parameters in models estimated within each of the treatment groups indicated that Figure series should load also the factor defined by the 2 verbal half-tests. With that parameter freed but the parameters being constrained to be equal in the groups a χ^2 of 40.3 was found, which with 38 degrees of freedom corresponds to a p-value of .37. This is here considered being an acceptable fit.

The aptitude variables thus fit a 3-factor model close to the hypothesized one. Following Cattell (1971) and Snow (1977) the 3 factor were labelled G_c (crystallized ability) G_f (fluid-analytic ability) and G_v (spatial-visualization ability), respectively.

But in this model there is a high correlation between G_f and G_v, the correlation in both groups being around .75. Such a correlation certainly is too high to give the statistical test of ATI effects any reasonable power with the sample sizes used. However, the variance represented by the G_v factor can be represented as a covariance between the errors of measurement for Flags and Figures instead. Therefore an alternative measurement model was defined in which the factors G_f and G_v were collapsed into one, G_{fv}. This model is shown in Figure 7.

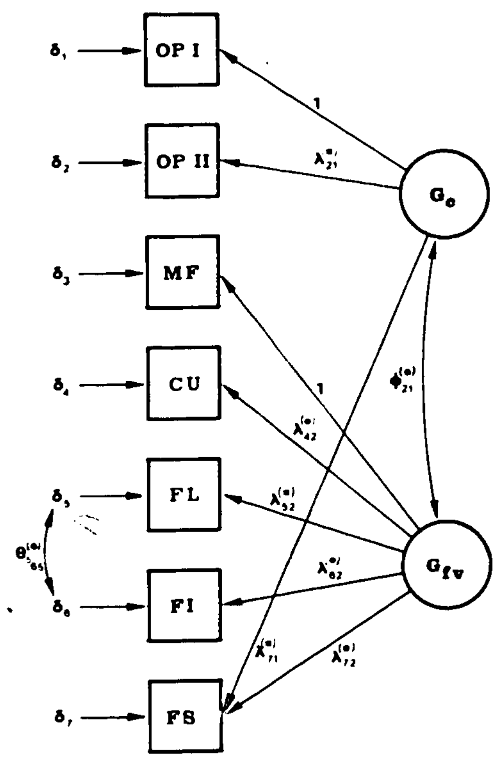

Figure 7 The 2-factor model for the aptitude variables in the Gustafsson (1979a) study

The following specification was used:

$$\underset{\sim}{\Lambda}_x^{(*)} = \begin{bmatrix} 1 & 0 \\ \lambda_{21}^{(*)} & 0 \\ 0 & 1 \\ 0 & \lambda_{42}^{(*)} \\ 0 & \lambda_{52}^{(*)} \\ 0 & \lambda_{62}^{(*)} \\ \lambda_{71}^{(*)} & \lambda_{72}^{(*)} \end{bmatrix}, \qquad \underset{\sim}{\Phi}^{(\ )} = \begin{bmatrix} \phi_{11}^{(*)} & \\ \phi_{21}^{(*)} & \phi_{22}^{(*)} \end{bmatrix},$$

$$\underset{\sim}{\Theta}_\delta^{(*)} = \begin{bmatrix} \theta_{\delta_{11}}^{(*)} & & & & & & \\ 0 & \theta_{\delta_{22}}^{(*)} & & & & & \\ 0 & 0 & \theta_{\delta_{33}}^{(*)} & & & & \\ 0 & 0 & 0 & \theta_{\delta_{44}}^{(*)} & & & \\ 0 & 0 & 0 & 0 & \theta_{\delta_{55}}^{(*)} & & \\ 0 & 0 & 0 & 0 & \theta_{\delta_{65}}^{(*)} & \theta_{\delta_{66}}^{(*)} & \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\delta_{77}}^{(*)} \end{bmatrix}.$$

A very good fit was found for this model ($\chi^2=40.9$, df=39, $p<.39$), and the covariance element in $\underset{\sim}{\Theta}_\delta^{(*)}$ was statistically highly significant. A moderate correlation of about .50 between G_c and G_{fv} was found.

This 2-factor model has the advantage that reasonable magnitudes of the standard errors of estimates of $\underset{\sim}{\Gamma}$ can be expected. But it has, of course, the disadvantage that if there is a different pattern of results for G_f and G_v this will not be detected, even at the descriptive level. Therefore we certainly cannot recommend that measurement models are developed with an eye only on the correlation between the ξ-variables; the substantive question under investigation is of course more important, and it does seem that ATI researchers must allow themselves to interpret even nominally non-significant effects (cf. Cronbach, 1975).

3. Multivariate regression on latent aptitude variables

So far we have only considered the case when there is just 1 y-variable. But LISREL can handle any number of outcome variables, and in the simplest case these are treated as separate observed outcome variables in a model corresponding to multivariate regression analysis. We will illustrate such a model through bringing in also the outcome variables in the Gustafsson (1979a) study.

EXAMPLE 4: Multivariate regression on 2 latent aptitude variables

The purpose of the study reported by Gustafsson (1979a) was to study suppression of visualization by reading (Brooks, 1967) and in particular if pupils with different aptitudes are differentially affected by such suppression. As has already been mentioned, there were 2 treatment groups, one that read an unillustrated teaching material dealing with the heart and the circulation of blood (Reading), and one that listened to a tape recorded presentation of the same information (Listening). It was suspected that some parts of this material is better learned if visualization processes are relied upon, and the results obtained by Brooks (1967) suggest that the possibilities of doing that should be better in the Listening treatment than in the Reading treatment.

Immediately after the instruction the pupils took 2 post-tests, one verbal and one pictorial. Using procedures described by Gustafsson (1979a) 3 scales were derived from the items in the 2 post-tests: one consisting of verbal questions asking about the circulation of blood (CIRC-V), one consisting of pictorial items asking about the circulation of blood (CIRC-P), and one consisting of verbal items asking about terms and other information of a verbal kind (VERB).

The covariances for these 3 outcome scales, as well as their covariances with the aptitude variables are presented in Table 4.

Table 4

The covariance matrices for the outcome scales, and the covariances between aptitudes and outcomes in the Gustafsson (1979) study.

	Reading			Listening		
	VERB	CIRC-V	CIRC-P	VERB	CIRC-V	CIRC-P
VERB	3.407			3.163		
CIRC-V	1.768	2.659		1.006	2.111	
CIRC-P	1.089	1.422	2.146	.890	.584	2.086
Opposites I	3.443	2.968	1.755	2.209	.980	.701
Opposites II	3.669	3.063	2.043	2.188	.922	.749
Metal folding	6.546	5.317	4.092	3.166	1.789	1.501
Cubes	6.045	5.442	3.310	2.023	1.770	2.122
Flags	22.008	21.071	16.329	9.619	12.063	6.017
Figures	15.013	13.562	8.549	5.558	6.085	3.856
Figure series	5.862	4.415	2.624	2.923	1.856	1.978

To analyze the data the 3 outcome scales were used as observed outcome variables, and these were regressed on the latent aptitude variables in the 2-factor model. This model is shown in Figure 8 and those parameter matrices which were not specified in the preceeding example are given below:

$$\underset{\sim}{\Lambda}_y^{(*)} = \underset{\sim}{I}, \quad \underset{\sim}{\Theta}_\varepsilon^{(*)} = 0, \quad \underset{\sim}{B}^{(*)} = \underset{\sim}{I}, \quad \underset{\sim}{\Gamma}^{(i)} = \begin{bmatrix} \gamma_{21}^{(i)} & \gamma_{12}^{(i)} \\ \gamma_{21}^{(i)} & \gamma_{22}^{(i)} \\ \gamma_{31}^{(i)} & \gamma_{32}^{(i)} \end{bmatrix},$$

$$\Psi^{(i)} = \begin{bmatrix} \psi_{11}^{(i)} & & \\ \psi_{21}^{(i)} & \psi_{22}^{(i)} & \\ \psi_{31}^{(i)} & \psi_{32}^{(i)} & \psi_{33}^{(i)} \end{bmatrix}$$

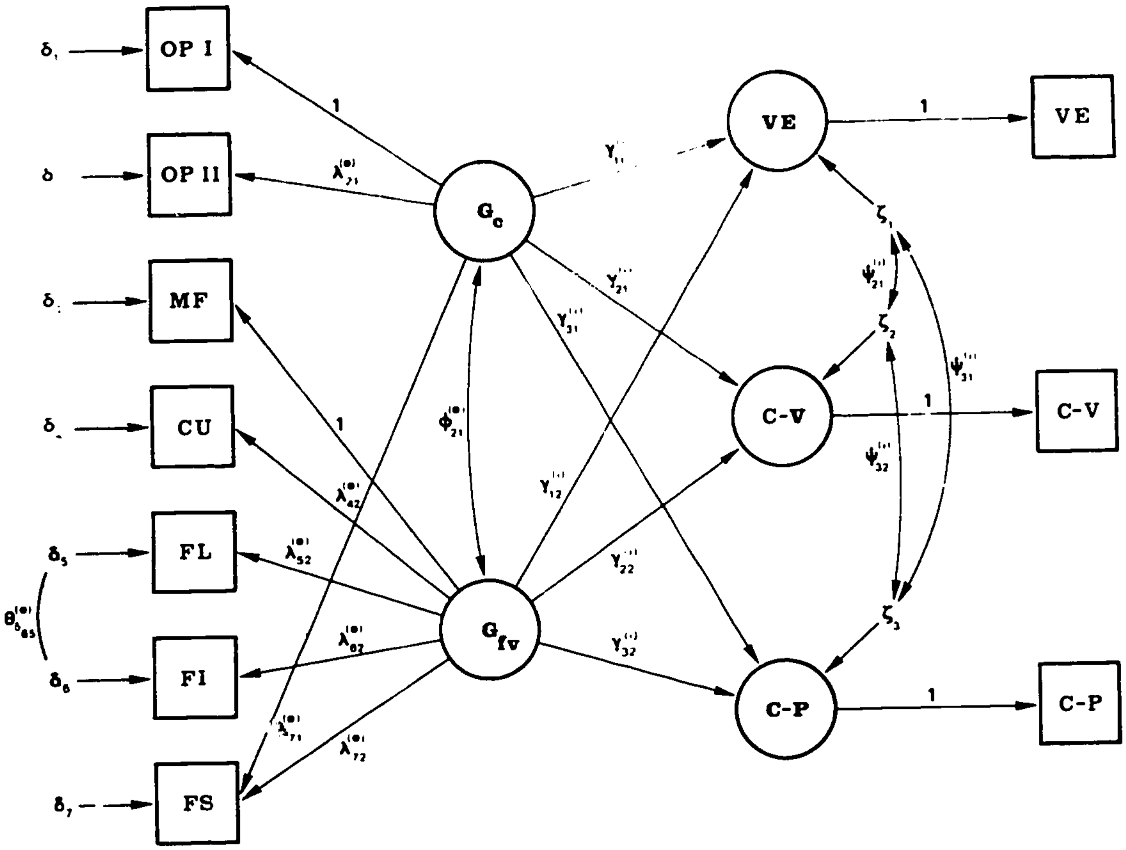

Figure 8 The model for the multivariate regression of the 3 observed outcome scales on the 2 latent aptitude variables in the Gustafsson (1979a) study.

Thus, in a multivariate regression analysis the $\underset{\sim}{\beta}$ matrix is considered being an identity matrix, and $\underset{\sim}{\Psi}$ must generally be taken to be a free symmetric matrix.

Table 5 presents the estimates of the $\underset{\sim}{\Gamma}^{(i)}$ matrices within the

Table 5

Estimates of the structural relations coefficients for the Gustafsson (1979a) study

	Latent aptitude variable					
	Reading	Listening	t	Reading	Listening	t
VERB	.33 (.06)	.28 (.07)	.54	.11 (.03)	.07 (.04)	.94
CIRC-V	.28 (.05)	.07 (.06)	2.59	.10 (.03)	.09 (.03)	.15
CIRC-P	.16 (.05)	.05 (.06)	1.40	.07 (.03)	.08 (.03)	-.24

2 treatment groups. The test of equality of the $\underset{\sim}{\Gamma}^{(i)}$coefficients within the treatments gave $\chi^2=12.2$, df=6, p<.06, so the interaction has a borderline significance. In Table 5 the results from pair-wise t-tests of the equality of each of the $\underset{\sim}{\Gamma}^{(i)}$-coefficients have been entered as well. The t-values have been computed according to the following formula:

$$t = \frac{\gamma^{(1)}_{y\xi} - \gamma^{(2)}_{y\xi}}{\sqrt{SE^2_{\gamma_{y\xi}}(1) + SE^2_{\gamma_{y\xi}}(2) - 2COV\ (SE_{\gamma_{y\xi}}(1)\ SE_{\gamma_{y\xi}}(2))}}$$

When the measurement model for the aptitude variables has been constrained to be equal in the 2 groups there is a slight correlation between the estimates of the $\underset{\sim}{\Gamma}$-coefficients, which must be taken into account.

The t-tests indicate that the interaction is accounted for by differences between the treatment groups with respect to the relation between G_c and CIRC-V, and to some extent also CIRC-P, there being a higher relation in the Reading group than in the Listening group between these variables.

Contrary to expectations no interaction is found with G_{fv}; it had been expected that a higher relation would be found between spatial ability and the spatial types of outcome variables in the Listening group than in the Reading group. But it is of course conceivable that the component parts of G_{fv} do give different patterns of results. Indeed, using the 3-factor model for the aptitude variables a higher relation between G_v and CIRC-V was found in the Listening treatment and also a higher relation between G_f and CIRC-P in that treatment. But in spite of the fact that rather large differences were found, they were far from significant, owing to the large standard errors of estimates caused by the high correlation between G_f and G_v. For a discussion about the results the reader is referred to Gustafsson (1979a).

4. Causal orderings among the outcome variables

In multivariate regression models no ordering is assumed between the outcome variables. But when the variables measuring outcome have been given at different points in time, to assess achievement and retention for example, there is such an ordering which should be reflected in the model. LISREL allows formulation of such models, in which a causal structure is specified not only between aptitude and outcome variables, but also among the outcome variables. The analysis thus performed corresponds to path analysis.

But in this kind of models errors of measurement in the outcome variables enter bias into the estimates, which can be taken into account either through correction for attenuation or through specifying a model with latent outcome variables. We will consider an example where only observed outcome variables are available, but where latent aptitude variables can be employed.

Example 5: Latent aptitude variables and observed outcome variables with a causal ordering

Out example consists of a reanalysis of an ATI study reported by Gagné and Gropper (1965; cf. Cronbach and Snow, 1977, pp. 96-99 and 266-273), which investigated differential effects of visual and verbal presentations. The design used was rather elaborate, and it included several kinds of measures of performance. Here we will use just a sub-set of variables; thus no full account of the study will be given.

The main instruction consisted in 7 self-paced programmed lessons on mechanical advantage, using a verbal presentation format. All subjects received this instruction. The treatments contrasted consisted of fixed-paced introductions with information presented either visually or verbally, and this introductory lesson functioned as an "advance organizer".There was also a control group which had no introductory lesson but this group will not be included in the reanalysis. But even before the experimental treatments, the subjects had instruction on concepts basic to the main program, in order to assure a common level of

prior knowledge.

The treatment groups were quite small, there being 46 and 42 subjects in the visual and verbal groups, respectively.

The study included 4 aptitude variables: Otis IQ (IQ), DAT Verbal reasoning (VR), DAT Abstract Reasoning (AR) and DAT Space Relations (SR). Before the treatments the subjects also were given a pre-test of mechanical advantage (PRE-ACH).

Achievement (ACH) was measured immediately after the main lessons, and retention (RET) was measured 4 weeks later. The same test was used on both occasions.

The covariance matrices are presented in Table 6

Table 6

The covariance matrices for the verbal and visual groups in the Gagné and Gropper (1965) study

Visual group:

	IQ	VR	AR	SR	PRE-ACH	ACH	RET
IQ	50.211						
VR	14.935	27.626					
AR	7.281	10.347	51.624				
SR	27.008	22.269	75.375	342.324			
PRE-ACH	-2.916	3.167	13.991	37.523	53.993		
ACH	10.198	19.098	23.537	53.842	20.565	63.218	
PET	9.700	16.269	24.620	47.589	23.320	57.193	76.021

Verbal group:

	IQ	VR	AR	SR	PRE-ACH	ACH	RET
IQ	55.383						
VR	31.168	55.339					
AR	33.934	31.169	64.674				
SR	63.570	40.860	82.057	382.085			
PRE-ACH	21.634	10.840	16.835	50.317	54.716		
ACH	18.420	28.721	29.438	54.174	24.826	64.160	
RET	15.804	34.870	16.920	17.790	6.732	44.131	95.766

The time ordering inherent in the design makes it possible to specify causal relationships between PRE-ACH, ACH and RET, and the outcome variables can then be "regressed" on the aptitude variables, which here can be estimated as latent aptitude variables.

A 1-factor model was first tried for the aptitude variables, but even with the small samples used, the fit of this model turned out to be poor (χ^2=22.1, df=6, p<.003).

It is reasonable to assume that a 2-factor model is needed to represent the aptitude variables, with IQ and VR measuring one factor (G_c) and AR and SR measuring the other factor (G_f). This model did show a good fit, even though the goodness-of-fit test cannot be taken seriously with the small groups available. In the sequel we will therefore stress the descriptive pattern of results.

In developing the measurement model for the aptitude variables it was noted that the correlation between the latent variables differed greatly in the treatment groups: in the Visual group it was .38 and in the Verbal group the correlation was .84.

The full LISREL model with G_c and G_f as latent aptitude variables is shown in Figure 9.

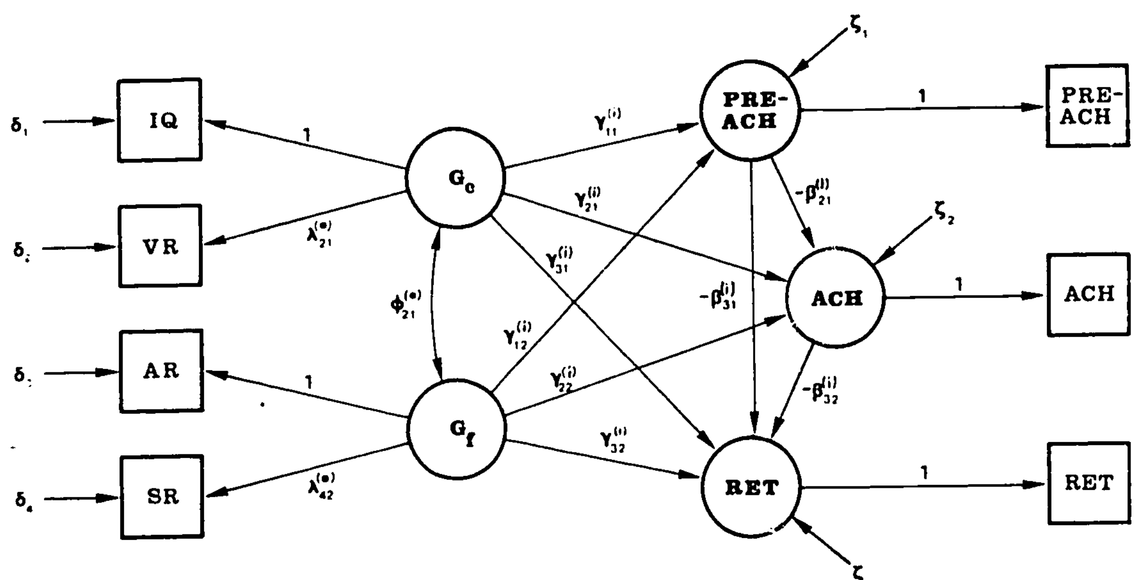

Figure 9 The model used for the Gagné and Gropper (1965) study.

and the following matrix specifications was used to estimate the model:

$$\Lambda_x^{(*)} = \begin{bmatrix} 1 & 0 \\ \lambda_{21}^{(*)} & 0 \\ 0 & 1 \\ 0 & \lambda_{42}^{(*)} \end{bmatrix}, \quad \Phi^{(*)} = \begin{bmatrix} \phi_{11}^{(*)} & \\ \phi_{21}^{(*)} & \phi_{22}^{(*)} \end{bmatrix},$$

$$\Theta_\delta^{(*)} = \text{diag } (\Theta_{\delta_{11}}^{(*)}, \Theta_{\delta_{22}}^{(*)}, \Theta_{\delta_{33}}^{(*)}, \Theta_{\delta_{44}}^{(*)}), \quad \Gamma^{(i)} = \begin{bmatrix} \gamma_{11}^{(i)} & \gamma_{12}^{(i)} \\ \gamma_{21}^{(i)} & \gamma_{22}^{(i)} \\ \gamma_{31}^{(i)} & \gamma_{32}^{(i)} \end{bmatrix},$$

$$\beta^{(i)} = \begin{bmatrix} 1 & & \\ -\beta_{21}^{(i)} & 1 & \\ -\beta_{31}^{(i)} & -\beta_{32}^{(i)} & 1 \end{bmatrix}, \quad \Lambda_y^{(*)} = I, \quad \Theta_\varepsilon^{(*)} = 0,$$

$$\Psi^{(i)} = (\psi_{11}^{(i)}, \psi_{22}^{(i)}, \psi_{33}^{(i)})$$

The causal relations among the y-variables are specified in β , which can be seen in the explicity formulated structural equations:

PRE-ACH= $\gamma_{11}G_v + \gamma_{12}G_f + \zeta_1$,

ACH= β_{21}PRE-ACH + $\gamma_{21}G_v + \gamma_{22}G_f + \zeta_2$,

RET= β_{22}ACH + β_{21}PRE-ACH + $\gamma_{31}G_v + \gamma_{32}G_f + \zeta_3$

The estimates of the $\Gamma^{(i)}$ and $\beta^{(i)}$ matrices are presented in Table 7. The relations between the latent aptitude variables and PRE-ACH and ACH differ only little between treatments, there being in both treatments relatively high relationships between G_f and PRE-ACH and between G_c and ACH. For RET, however, there is in the Visual group no relationships with any of the latent aptitude variables, while in the Verbal group there is a rather strong positive relationship between RET and G_c, and

Table 7

Estimates of the structural relations coefficients for the Gagné and Gropper (1965) study

	$\Gamma^{(i)}$				$B^{(i)}$			
	G_c		G_f		ACH		RET	
	Visual	Verbal	Visual	Verbal	Visual	Verbal	Visual	Verbal
PRE-ACH	-.28(.43)	.08(.43)	.52(.30)	.39(.30)	.26(.15)	.26(.15)	.08(.11)	-.19(.19)
ACH	.89(.43)	.51(.41)	.31(.30	.28(.29)			.86(.13)	.66(.21)
RET	-.07(.33)	1.07(.55)	.08(.22)	-.55(.37)				

also a rather strong negative relationship between RET and G_f. In the Verbal treatment there is also a lower relationship between ACH and RET than there is in the Visual treatment.

Cronbach and Snow (1977, pp. 266-273) have presented another reanalysis of this study and that reanalysis gave very much the same results as those found here. In the Cronbach and Snow reanalysis 2 weighted composites of the 4 aptitude variables were used, however, and in that analysis it could not easily be seen that there is a negative partial relationship between G_f and RET in the Verbal group.

As was also pointed out by Cronbach and Snow the results are, however, affected by errors of measurement in the outcome variables. No estimates of reliability are presented by Gagné and Gropper but to illustrate the effects of errors of measurement in the outcome variables, we have corrected for attenuation using an arbitrary estimate of .90 as the reliability of the 3 outcome variables.

Correction for attenuation in the outcome variables is of course done in exactly the same way as correction for attenuation in the aptitude variables, and we present here only the fixed parameters in the $\Theta_\varepsilon^{(i)}$ matrices:

$\Theta_{\varepsilon}^{(VI)} = (5.399, 6.322, 7{,}602)$, $\Theta_{\varepsilon}^{(VE)} = (5.472, 6.416, 9.577)$

The estimates of the $\Gamma^{(i)}$ and $\beta^{(i)}$ matrices, after correction for attenuation are presented in Table 8. Among the Γ coefficients only the coefficient for the relation between G_c and RET in the Visual group is affected to any appreciable extent. The coefficients for the relation between ACH and RET are the ones most affected, but the difference between the treatment groups of course remains. Thus, the unattenuated coefficients allow very much the same conclusions as the attenuated ones.

Table 8

Estimates of the structural relations coefficients for the Cagné and Gropper (1965) study, using correction of attenuation for the outcome variables

	$\Gamma^{(i)}$				$\beta^{(i)}$			
	G_c		G_f		ACH		RET	
	Visual	Verbal	Visual	Verbal	Visual	Verbal	Visual	Verbal
PRE-ACH	-.28(.43)	.08(.43)	.52(.30)	.39(.30)	.29(.17)	.30(.17)	.04(.14)	-.25(.22)
ACH	.90(.43)	.51(.41)	.30(.30)	.26(.29)			1.05(.17)	.81(.25)
RET	.23(.37)	1.00(.54)	.02(.23)	-.58(.38)				

Correction for attenuation is not the optimal solution of the problem of errors of measurement in the outcome variables; if possible a measurement model with latent variables should be used for the outcome variables.

To conclude , there is a tendency towards interaction with respect to RET. However, the treatment groups were small, and probably more importantly, there was a large difference between the correlation for the latent aptitude variables in the treatments. Thus, had for example G_f been used as an outcome variable, and G_c as an aptitude variable, at least as strong interactions as those found with the outcome variables proper would have been found, which should temper attempts to interpret the results in

this study until they have been replicated.

5. Structural relations between latent aptitude and latent outcome variables

Especially when there is a causal ordering of the outcome variables it is essential that latent rather than observed outcome variables are studied. But also when many outcome variables have been measured at the same time, it may be desirable to reduce these to a smaller set of η-variables.

Using a concrete example we will illustrate how such models with both latent aptitude and latent outcome variables can be formulated.

EXAMPLE 6: Latent aptitude variables/latent outcome variables

Our example consists of a reanalysis of a study presented by Behr (1967) and for a full account of the reanalysis the reader is referred to Gustafsson and Lindström (1978).

Behr investigated the same hypothesis as did Gagné and Gropper, i.e. that tests of verbal ability are more highly correlated with achievement in a verbal treatment than in a figural treatment and that tests of spatial (or figural) ability are more highly related to achievement in a figural treatment than in a verbal treatment. The subject matter taught was modulus seven arithmetic and there were 2 treatment groups: in a verbal-symbolic treatment (VS) subjects studied a programmed teaching material using algebraic symbols complemtented with verbal information; in a figural-symbolic (FS) treatment figural information was added.

The aptitude variables were selected to correspond to cells in the Guilford (1967) "Structure-of-Intellect (SI)" model. Of the tests, 6 had a figural content (-F-), 5 had a sematic tontent (-M-) and 2 had a symbolic (-S-) content. In addition there was a test called Integration which was not directly classifiable in the SI structure. There are reasons, however, which motivate this test being regarded an -F- test (cf. Gustafsson & Lindström, 1978, p.4). The aptitude variables are presented in Table 9.

Table 9

The aptitude variables in the Behr (1967) study

Test	SI-cell
Gestalt Completion Test	CFU
Figure Classification	CFC
Figure Matrix	CFR
Paper Folding Test	CFT
Map Memory	MFU
Object Memory	MFS
Wide Range Vocabulary Test	CMU
Word Classification	CMC
Verbal Analogies	CMR
Memory for Word Meanings	MMU
Sentence Completion	MMR
Object-Number Test	MSR
Addition/Subtraction Test	MSI
Following Direction Test	(INT)

Three criterion measures were determined: Time used to study the program (TP), a Learning Test (LT) score and a Retention Test (RT) score. The LT was administered 2 days after the instruction and the RT 2 weeks after the instruction. The LT and RT were parallel forms and consisted each of 5 parts. From these parts 2 subtest scores were derived: LA and RA which were speed tests of modulus seven addition and subtraction, and LB and RB which assessed understanding of structural properties of the modulus seven system.

Subjects in the study were prospective elementary school teachers, there being 120 and 109 subjects in the VS and FS groups, respectively. The subjects were randomly assigned to treatments.

In the reanalysis the measurement models for the aptitude and outcome variables were first developed separetely, and they were then fitted together through the structural relations equations.

The measurement model for the aptitude variables included the 7 -F- tests and 5 -M- tests, which were assigned to one factor each in a 2-factor model. The goodness-of-fit test showed a borderline significance (χ^2=128.0, df=106, p<.07) when the parameters of the measurement model were allowed to be different within the treatment groups; constraing them to be equal the test of fit resulted in a p-value of .05. The fit could be better but it did seem difficult to develop any useful measurement model with a better fit, so this model was used, along with another measurement model with 1 latent variable only.

The latent variables in the 2-factor model will be referred to as factors of verbal and figural ability. Unfortunately there was a high correlation of around .80 between these factors in both groups, and we have already seen how such high a correlation effectively precludes discovery of significant ATI effects. Several attempts were made to find other measurement models with a lower correlation between the latent aptitude variables. These attempts were unsuccessful, however, and it does seem that the analysis of the aptitude variables shows that there simply is not much information about the verbal/figural ability distinction in this sample. This is almost certainly due to the facts that the majority, if not all the subjects in the sample must have been female, and that tests of figural ability were used in which the items can successfully be solved by verbal and non-verbal reasoning processes. There are strong indications that females in particular resort to such strategies on figural tests whenever possible (Gustafsson, 1976, ch 6).

The measurement model for the outcome variables included the subtest scores (LA, LB, RA, and RB) from the criterion tests. With 4 observed variables it is in LISREL possible to define a model with 2 latent variables, and still there is 1 degree of freedom left to test goodness of fit. But it should be

possible to define 2 potentially meaningful 2-factor solutions: either one with a learning factor defined by LA and LB, and a retention factor defined by RA and RB, or a model with a computational speed factor defined by LA and RA, and a factor reflecting understanding of the structural properties of the modulus seven system, defined by LB and RB.

Both these 2-factor models were estimated and tested, and in both cases a poor fit was found, with the fit being worse for the learning/retention factors than for the speed/understanding factors. This indicates that in fact all 4 latent variables are needed to account for the structure among the 4 observed variables. But again there is the possibility of invoking covariances between the errors of measurement, and in that way representing systematic variance which would otherwise call for additional latent variables.

Thus, when the computational speed/understanding factors are postulated, for example, the possible effects on the correlation between these 2 latent variables of administration of the tests at 2 common occasions, can be taken into account if $\Theta_{\varepsilon_{LA,LB}}$ and $\Theta_{\varepsilon_{RA,RB}}$ are allowed to be greater than zero.

Quite obviously, however, such a model cannot be estimated within the framework of the measurement model for the y-variables only -- with 4 observed variables and 2 latent variables there is only 1 degree of freedom left. However, as was shown by Jöreskog & Sörbom, (1978, pp. 22-28) it is possible to estimate such a model when ξ variables are added.

The full LISREL model for the computational speed/understanding factors and for the verbal/figural aptitude factors is shown in Figure 10, for one of the treatment groups. The LISREL model for the learning/retention factors is quite similar to this one, the only differences being that there is a unidirectional influence of learning on retention, and that LA and RA in that model have correlated errors of measurement, as have LB and RB.

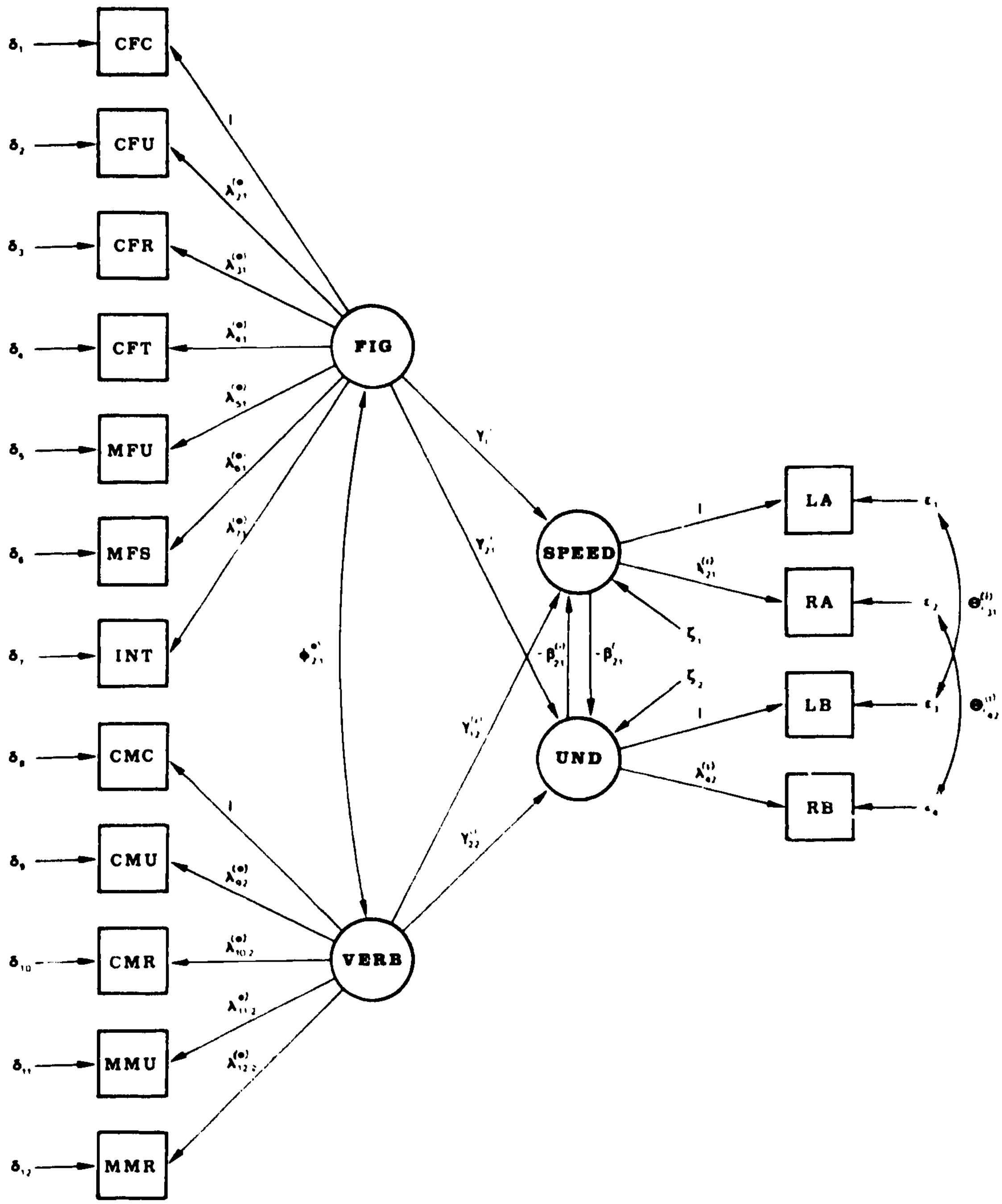

Figure 10 The full LISREL model for the speed/understanding outcome factors in the Behr (1967) study.

To estimate the model containing the computational speed/understanding factors as outcome variables, the following specification was used:

$$\underset{\sim}{\Lambda}_x^{(*)} = \begin{bmatrix} 1 & 0 \\ \lambda_{21}^{(*)} & 0 \\ \lambda_{31}^{(*)} & 0 \\ \lambda_{41}^{(*)} & 0 \\ \lambda_{51}^{(*)} & 0 \\ \lambda_{61}^{(*)} & 0 \\ \lambda_{71}^{(*)} & 0 \\ 0 & 1 \\ 0 & \lambda_{92}^{(*)} \\ 0 & \lambda_{102}^{(*)} \\ 0 & \lambda_{112}^{(*)} \\ 0 & \lambda_{122}^{(*)} \end{bmatrix}, \quad \underset{\sim}{\Theta}_\delta^{(*)} = \text{diag}\ (\Theta_{\delta_{11}}^{(*)}, \ldots, \Theta_{\delta_{1212}}^{(*)}),$$

$$\underset{\sim}{\Phi}^{(*)} = \begin{bmatrix} \phi_{11}^{(*)} & \\ \phi_{21}^{(*)} & \phi_{22}^{(*)} \end{bmatrix} \qquad \beta^{(i)} = \begin{bmatrix} 1.0 & -\beta_{21}^{(i)} \\ -\beta_{21}^{(i)} & 1.0 \end{bmatrix}$$

$$\underset{\sim}{\Lambda}_y^{(i)} = \begin{bmatrix} 1 & 0 \\ \lambda_{21}^{(i)} & 0 \\ 0 & 1 \\ 0 & \lambda_{42}^{(i)} \end{bmatrix}, \quad \underset{\sim}{\Theta}_\varepsilon^{(i)} = \begin{bmatrix} \theta_{\varepsilon_{11}}^{(i)} & & & \\ 0 & \theta_{\varepsilon_{22}}^{(i)} & & \\ \theta_{\varepsilon_{31}}^{(i)} & 0 & \theta_{\varepsilon_{33}}^{(i)} & \\ 0 & \theta_{\varepsilon_{42}}^{(i)} & 0 & \theta_{\varepsilon_{44}}^{(i)} \end{bmatrix},$$

$$\underset{\sim}{\Gamma}^{(i)} = \begin{bmatrix} \gamma_{11}^{(i)} & \gamma_{12}^{(i)} \\ \gamma_{21}^{(i)} & \gamma_{22}^{(i)} \end{bmatrix}, \quad \underset{\sim}{\Psi}^{(i)} = \text{diag}\ (\psi_{11}^{(i)}, \psi_{22}^{(i)})$$

The 2 free parameters in the $\underset{\sim}{\beta}$-matrix have been constrained to be equal which is indicated by use of the same index to refer to them within the matrix. The parameters of the measurement model for the aptitude variables were constrained to be equal in the treatment groups. The parameters of the measurement model for the outcome variables were not constrained to be equal in the treatment groups, even though it should perhaps have been desirable to constrain $\underset{\sim}{\Lambda}_y$ as well.

The model had a far from perfect fit (χ^2=272.0, df=217, p<.01), but we already know that the most important source of lack of fit is the measurement model for the aptitude variables.

The t-values of the estimates of the covariance for the errors of measurement of LA and LB was 1.44 and 1.39 in the VS and FS groups respectively; for RA and RB the corresponding figures were 3.15 and .20. Thus, only in the VS group a significant covariance is found.

The test of equality of the $\underset{\sim}{\Gamma}^{(i)}$ matrices gave χ^2=3.2, df=4, p<.53, so the interaction is not significant, as can hardly be expected with as high a correlation between the latent aptitude variables as is present in these data.

The coefficients are presented in Table 10. Descriptively

Table 10

Estimates of the structural relation coefficients for the speed/understanding outcome variables in the Behr (1967) study.

	Speed			Understanding		
	VS	FS	t	VS	FS	t
Verbal	.24(.85)	-1.01(.75)	1.10	.63(.70)	1.20(.85)	-.52
Figural	.32(.23)	.44(.20)	-.39	.37(.19)	.14(.23)	.77

Standard errors are shown in parentheses.

The t-values refer to approximate tests of equality of the within-treatment coefficients.

there are some differences between the $\underset{\sim}{\Gamma}^{(i)}$ coefficients, and for a discussion about possible interpretations of the tendency towards interaction, the reader is referred to Gustafsson and Lindström (1978).

Neither with respect to the learning/retention outcome factors any significant interaction was found and for these outcome there were not even descriptively any differences between the $\underset{\sim}{\Gamma}^{(i)}$ coefficients worth noting. Bur the relation between learning and retention was stronger within the FS treatment than within the VS treatment, the β coefficients being .76 and .62, respectively. Thus, in this study just as in the Gagné and Cropper study there is a tendency for retention to be less related to learning in a verbal than in a visual/figural treatment. Such differential effects of treatment have not previously been given much attention in ATI research because the outcome variables have been treated on at a time in separate MR analyses.

In the original analysis of the data Behr (1967) examined the within-treatment regressions of one outcome variable at a time on one aptitude variable at a time. A handful of significant interactions were found, with CMU, MMR and CMC entering interactions with one or more outcomes. It is hard to find any reason, however, why these verbal tests and not the others should enter into interactions.

6. Formulating LISREL models including intercept parameters

All the LISREL models formulated hitherto are incomplete as models for ATI effects in the sense that they do not include the intercept parameters. Without estimates of these parameters it is impossible to determine whether an interaction is ordinal or disordinal, for example.

But, as has been shown by Sörbom (1978, 1979) it is possible to formulate LISREL models which do allow estimation and testing of these parameters as well. Using the generated data constructed in Example 1 and analyzed in Example 2, we will illustrate the procedure.

EXAMPLE 7: Univariate regression on latent aptitude variables, including the intercept

The structural equation models used in Example 2 can in explicit notation be written:

$$y^{(1)} = \gamma_1^{(1)}\xi_1 + \gamma_2^{(1)}\xi_2 + \zeta;$$

$$y^{(2)} = \gamma_1^{(2)}\xi_1 + \gamma_2^{(2)}\xi_2 + \zeta.$$

Here, however, we want to estimate instead the structural equations:

$$y^{(1)} = \alpha^{(1)} + \gamma_1^{(1)}\xi_1 + \gamma_1^{(1)}\xi_2 + \zeta;$$

$$y^{(2)} = \alpha^{(2)} + \gamma_1^{(2)}\xi_1 + \gamma_2^{(2)}\xi_2 + \zeta,$$

where $\alpha^{(i)}$ is used to denote the intercept parameter.

In generating the data all the variables were supposed to have a zero mean. The following sample mean vectors were observed:

$$\bar{\underset{\sim}{x}}^{(1)} = (-.090,\ .061,\ -.061,\ -.137,\ -.005)$$

$$\bar{\underset{\sim}{x}}^{(2)} = (-.099,\ -.086,\ -.148,\ -.055,\ -.260)$$

with the y variable given first, and then the 4 x-variables.

To take into account the vectors of means, along with the covariances, the matrix of moments around zero is analyzed. The moment matrix is computed automatically by the program if the mean vector and the covariance matrix are both supplied.

A dummy variable, which for all the persons has the value 1, must be added, however. The dummy variable is the only x-variable in the model and it is treated as a fixed variable. All the other variables are treated as y-variables, with the pattern of relations between the variables being specified in

the $\underset{\sim}{\beta}$-matrix. Another η-variable must also be included, with which the means of the variables are represented.

For our example the following specification was used.

$$\underset{\sim}{\Lambda}^{(*)} = (1),\ \underset{\sim}{\Theta}_{\delta}^{(*)} = (0),\ \underset{\sim}{\Phi}^{(*)} = (1),\ \underset{\sim}{\Lambda}^{(*)} = \begin{bmatrix} 1 & 0 & 0 & \lambda_{14}^{(*)} \\ 0 & 1 & 0 & \lambda_{24}^{(*)} \\ 0 & \lambda_{32}^{(*)} & 0 & \lambda_{34}^{(*)} \\ 0 & 0 & 1 & \lambda_{44}^{(*)} \\ 0 & 0 & \lambda_{53}^{(*)} & \lambda_{54}^{(*)} \end{bmatrix},$$

$$\underset{\sim}{\beta}^{(i)} = \begin{bmatrix} 1 & -\beta_{12}^{(i)} & -\beta_{13}^{(i)} & 0 \\ 0 & 1 & -\beta_{23}^{(*)} & 0 \\ 0 & -\beta_{23}^{(*)} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},\ \underset{\sim}{\Gamma}^{(1)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix},\ \underset{\sim}{\Gamma}^{(2)} = \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \\ \gamma_{31} \\ 1 \end{bmatrix},$$

$$\Theta_{\varepsilon}^{(*)} = \text{diag}\ (0,\ \Theta_{\varepsilon_{22}}^{(*)},\ \Theta_{\varepsilon_{33}}^{(*)},\ \Theta_{\varepsilon_{44}}^{(*)},\ \Theta_{\varepsilon_{55}}^{(*)}),$$

$$\underset{\sim}{\Psi}^{(*)} = (\psi_{11}^{(i)},\ \psi_{22}^{(i)},\ \psi_{33}^{(i)},\ 0).$$

The $\gamma_{11}^{(i)}$ parameters here represent the intercept parameters and the other elements of the $\underset{\sim}{\Gamma}^{(i)}$ matrices represent the means on the latent aptitude variables. It is generally only possible to study differences between the treatment groups, with respect to the means on the latent variables and the $\gamma_{11}^{(1)}$-$\gamma_{31}^{(1)}$ parameters have been specified to be equal to zero, while the corresponding parameters in the other treatment groups are free parameters.

The parameters in this model which correspond to those estimated in Example 2 were quite close, and the small differences which could be observed are due to the fact that here the information in the mean vectors is also taken into account.

The estimates of the free parameters in $\underset{\sim}{\Gamma}^{(2)}$ were .08, -.04 and -.12, respectively and we can write the full structural equation models as:

$$y^{(1)} = 0 + .68\xi_1 + .18\xi_2$$

$$y^{(2)} = .08 + .05\xi_1 + .61\xi_2$$

This method of estimating the intercept parameters can be used with latent as well as with observed outcome variables, and it generalizes to any number of outcome variables, in which case a vector of intercept parameters is of course estimated for each group.

7. LISREL models for multilevel data

In the models considered so far the individual subjects' scores form the basis for estimation of parameters and testing goodness-of-fit. But in ATI research classes rather than pupils are often sampled, and the instructional process most often takes place with the pupils organized into classes; thus the pupils can often not be considered independent units of observation.

Having analyzed the possible consequences of treating such hierarchically nested observations as individual observations Cronbach and Snow (1977) were forced to make a "radical reappraisal of the ATI model" (p. 99; cf Cronbach, 1976b),asserting the necessity of separating within-class and between-class components of ATI effects. They showed that ATI effects may arise not only through pupils'differential response to treatments, but some processes may affect the class as a unit, and sometimes the pupil's relative standing in the class may be of functional importance.

To analyze ATI data at the between-class level Cronbach and Snow (1977) suggested that for each pupil the class means on the aptitude and outcome variables should be entered into the regression analysis, while for the within-class analysis it was suggested that the deviation scores between the pupils'

scores and their class means should be entered. It has been found that such analyses often give drastically different results at different levels (Cronbach, 1976b),even though the differences can often be accounted for by anomalies in the data (Cronbach & Webb, 1975; Gustafsson, 1978).

LISREL also offers great possibilities for conducting such multilevel analyses . It has been shown by Schmidt (1969) that maximum likelihood estimates can be derived of the within-class and between-class covariance matrices, and these can be parameterized in LISREL models, to allow separate estimates of parameters at the two levels, and also tests of the equality of structural relations, for example, at the two levels (Keesling, 1976, 1978). Such two-level analyses could in priciple be carried out for all the types of LISREL models we have considered here.

A great problem, of course, is that there in most studies tend to be few classes (or other higher level units) only, which precludes the possibility of obtaining any stable estimates at the class level. We would like to suggest, however, that in the least within-class analyses should be performed to guard against the possibility that results obtained in non-hierarchical analyses can in fact be accounted for by effects at the class level, which may be more or less artifactual.

EXAMPLE 8: Pooled within-class analyses contrasted with overall analyses

We will present, briefly, an example of such an analysis. The data analyzed were collected within a large scale observational study of the teaching process in 60 classes in grade 6 (Bredänge et al, 1971), which study also included measures of teacher and pupil personality and, for some of the classes, measures of learning outcome. Here only a small subset of variables will be analyzed, and for a full account of the results from reanalyses of these data the reader is referred to Gustafsson (1979d).

Both the pupils and the teachers were given the same personality questionnaire, the High School Personality Questionnaire (HSPQ, Cattell, Coan & Beloff, 1957). The items in that questionnaire have been reorgainzed, however, to measure three scales labelled

Introversion, Impulsivity and Stability, with the definition of these variables coming close to the corresponding Eysenckian concepts (Eysenck & Eysenck, 1969). Only the Extraversion/Introversion (E/I) variable will be studied here.

The 13 teachers scoring highest on the E/I scale were selected as an extreme group of I teachers, and the 12 teachers scoring lowest were selected as an extreme group of E teachers. In the I classes there were 289 pupils and in the E classes there were 254, who had a complete set of data.

Aptitude variables were the E/I scale, a vocabulary test of verbal ability (VA), and a test of nonverbal reasoning ability (RA). Since these variables reflect different latent variables, each variable was in the LISREL model entered as two half-tests, with the same λ coefficient for each.

Outcome variables were 5 sub-tests in a standardized mathematics ahcievement test, and a scale measuring attitude towards the teacher (Gustafsson, 1979b). The 5 mathematics sub-tests were in the measurement model for the dependent variables taken to reflect the same latent variable (MA), and the teacher attitude scale was entered as 2 half-test to make possible inference about the latent variable (TA).

The full LISREL model for these variables is shown in Figure 11, and there is no need here to present in detail the parameter specifications, since these can be derived from the figure and from the models treated earlier in this paper.

In one analysis no account was taken of the fact that the pupils were organized into classes, i.e. the covariance matrices computed from the raw scores were analyzed (Pooled). In another analysis the covariance matrices computed from the deviations between the pupils´scores and their respective class means were analyzed (Within), and the number of observations was taken to be the number of pupils minus the number of classes.

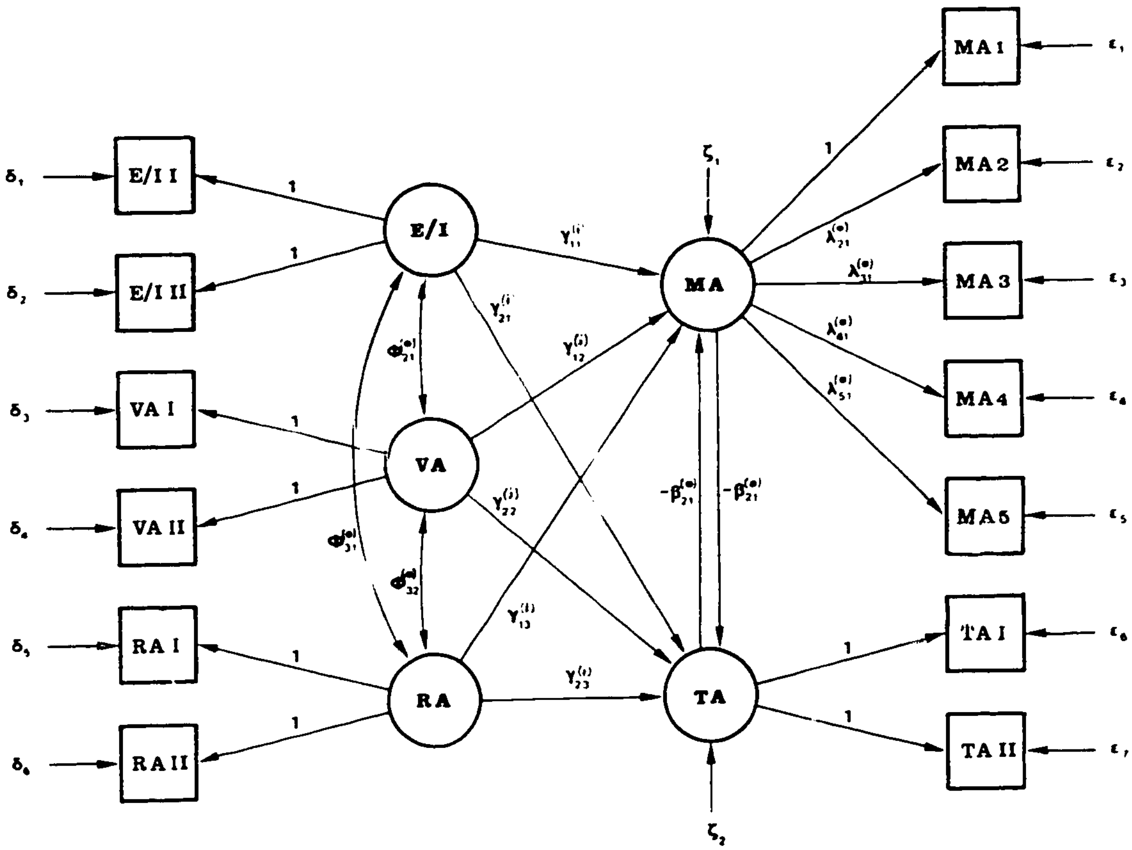

Figure 11 The model used to study interactions between teacher an pupil personality.

In the Pooled analysis, with $\Lambda_x^{(*)}$ and $\Phi^{(*)}$ constrained to be equal in the I and E groups, a rather poor fit was found, with a p-value of .0005. The main reason for this poor fit was that 1 factor only could not account for the 5 mathematics sub-tests. In the corresponding Within analysis, however, a p-value of .04 was found, which indicates that there may be differences between teachers with respect to which weight they place on different aspects of mathematics.

The estimates of the $\Gamma^{(i)}$ coefficients for the Pooled and the Within analyses are presented separately for the E and I groups in Table 11. The estimates in the Pooled and in the Within analyses generally come quite close, even though there also are some slight differences.

Table 11

Estimates of the structural relation coefficients in the Pooled and the Within analyses.

	Pooled				Within			
	MA		TA		MA		TA	
	E	I	E	I	E	I		I
VA	.91	.74	-.03	-.02	.88	.69	-.04	.06
RA	.68	.84	-.12	.03	.75	.81	-.04	-.08
E/I	-.18	.14	-.14	.26	-.19	.15	-.11	.28

The test of overall interaction gave in the Pooled analysis $\chi^2 = 8.3$, which with 6 degrees of freedom is not significant. In the Within analysis a somewhat higher, although non-significant, χ^2 of 9.6 was found. But it is reasonable to restrict here the test of interaction to the E/I scale, and in Table 11 it can be observed that there are important differences between the within-group coefficients for this variable. The test of interaction with Introversion alone gave in the Pooled analysis a χ^2 of 6.3, which with 2 degrees of freedom is significant; in the within analysis an even higher χ^2 of 7.6 was found.

Thus there is a significant interaction with introversion, such that in classes with an extravert teacher there is a negative relationship between this variable and achievement/attitude, while in classes with an introvert teacher there is a positive relationship. The fact that a somewhat stronger relationship is found in the Within analysis than in the Pooled analysis lends further credence to this conclusion, since the match/mismatch between teacher and pupils personality must take place within classes, and since the interaction cannot be accounted for as being due to a few classes having extreme means on one or more of the variables.

8. Discussion and conclusions

ATI studies are typically designed to test hypotheses formulated in terms of hypothetical constructs referring to unobservable aptitude and outcome variables, and as indicators of these variables several observed variables are often used. As we have tried to show in this paper, experiments with such a structure can advantageously be analyzed with LISREL.

Of course other methods can be used as well: In order to reduce several observed variables to few latent variables, component or factor analysis can be applied (cf.Cronbach & Snow, p. 39); to study relations between true variables rather than observed variables correction for attentuation may be employed; and to study relations between measurements with an intrinsic causal ordering path analysis can be used. But with LISREL it is possible to specify models including all these features, which results in a more parsimonious and often more efficient analysis.

LISREL also brings other advantages to ATI research. The necessity of formulating explicit measurement models for the aptitude variables makes it natural to investigate the similarity of the structure of the aptitude variables in the treatment groups, which only too seldom has been done. But when it is done, it is surprisingly often found that there are differences between the treatment groups with respect to the level and structure of the aptitude variables (cf. Cronbach & Webb, 1975; Cronbach & Snow, 1977, p. 38; Gustafsson, 1976, 1977). Such differences often result in spurious ATIs and they can of course also be suspected to conceal ATIs at times. Using LISREL it is easy to detect such differences and LISREL also allows some investigations into the possible effects of such differences on the estimates of the stuctural relations within treatments, through comparing the results when the same measurement model is used and when different measurement models are used.

The necessity in LISREL of formulating explicit models brings another advantage: It forces the researcher to a deeper penetration of the substantive problem, both in designing the study and in analyzing the data. In the long run this may prove to be one

of the greatest contributions of LISREL to ATI research.

The possiblity in LISREL to test the goodness-of-fit of a model or of a part of a model is another great advantage of the method. It must be stressed, however, that use of statistical tests to assess the fit of data to a model is fraught with several problems of which it is necessary to be aware (cf. Gustafsson, 1979c, pp. 27-28). For one thing, the test is a large sample test and when too small a sample is used, there is a risk that the test statistic does not have the distribution assumed. Another problem associated with small samples is that the power of the tests may be too low to detect even gross deviations from the model. But samples can be too large as well. This is because no model can ever be supposed to be perfectly fitted by data, so with large enough a sample any model would have to be discarded. The results from the goodness-of-fit test must therefore not be given any absolute interpretation, and modifications of models should be dictated more by substantive considerations, than by statistical ones.

So far we have stressed the advantages of LISREL over other methods for analyzing ATI data, but there are of course problems associated with this method of analysis as well.

In LISREL it is essentially assumed that data have a multivariate normal distribution. Only little is known about the robustness of the method against violations of this assumption, but it is probably fair to say that some of the advantages that LISREL has, are reduced when this assumption is not fulfilled.

It must also be pointed out that to achieve its full potential, LISREL requires more than one observed variable for each latent variable. As has already been pointed out this problem can be solved, however, if half-tests are entered. The method of half-tests is likely to be useful also when a group of tests do measure a common factor, but when there is also important specificity in some of the variables which is suspected to interact with treatments.

In our experience the standard errors of the estimates of the $\underset{\sim}{\Gamma}$ coefficients tend to be large in LISREL, which is a great problem.

Even in MR the power of the test of interaction generally is too low with the sample sizes which are feasible in ATI research (Cronbach & Snow, 1977).

One reason why the standard errors tend to be large in LISREL is that there often is an inverse relationship between the consistency and the variance of an estimator (Parenthetically it can be pointed out that even within regression analysis this inverse relationship has been studied,cf. Winer, 1978). But the standard errors are also functions of the size of the correlations between the aptitude variables, and another reason why LISREL tends to give higher standard errors than MR is that in MR these correlations are underestimated as a function of unreliability in the aptitude variables.

If statistically significant interactions are sought it is necessary to select the sample, the aptitude variables and the LISREL model so as to minimize the correlation between the latent aptitude variables. But as was pointed out by Cronbach and Snow (1977) the generally low power of any test of ATI effects makes it necessary to place lower weight on formal statistical tests, and to consider instead the descriptive patterns of results. Cronbach (1975) even claimed that: "The time has come to exorcise the null hypothesis. We cannot afford to pour our costly data down the drain whenever effects present in the sample 'fail to reach significance'"(p.124).

We strongly agree that less emphasis should be placed on statistical inference and that greater importance should be attached to description of effects in the sample. Even in such an approach it would seem that LISREL offers some advantages. For one thing, the latent aptitude variables can be supposed to be more or less invariant over different studies, which is important when the results from different studies are brought together. The description is also based on consistent estimates of the parameters and it is generally very parsimonious. Furthermore, when a descriptive approach is used it is important that an eye is kept on possible differences in the structure of the aptitude variables in the treatment groups, and, as has already been pointed out, this is easily done with LISREL.

For reasons of space we have generally not included results from MR analyses of the data analyzed with the different LISREL models. It is, however, our experience that the MR results may differ drastically from the LISREL results, as may of course be expected on the basis of the problems caused by errors of measurement only. We therefore want to conclude this paper by recommending that, to the extent it is possible, the available ATI data are reanalyzed with LISREL.

REFERENCES

Behr, M.J. (1967) A study of interactions between "structure-of-intellect" factors and two methods of presenting concepts of modulus seven artihmetic. Unpublished doctoral dissertation. Florida State University. Shorter version: Journal for Research in Mathematics Education, 1, 1970, 29-42.

Bergman, L.R. (1971) Some univariate models in studying change. Reports from the Psychological Laboratories, The University of Stockholm, Supplement 10.

Bredänge et al. (1971) Didaktisk processanalys. Presentaion av syften, uppläggning, undersökningsgrupper och mätinstrument samt några beskrivande data. (Didactical process analysis. Presentation of aims, design, samples and instruments, and some descriptive data.) Rapport nr 24 från Institutionen för praktisk pedagogik, Lärarhögskolan i Mölndal.

Brooks, L.R. (1967) The suppression of vizualization by reading. Quarterly Journal of Experimental Psychology, 19, 289-299.

Cattell,R.B. (1971) Abilities: their structure, growth and action. Boston: Houghton-Mifflin.

Cattell,R.B.,Coan,R.W., & Beloff, H. (1957) A re-examination of personality structure in late childhood and development of the High School Personality Questionnaire. Journal of Experimental Education, 27, 73-88.

Cronbach, L.J. (1957) The two disciplines of scientific psychology. American Psychologist, 12, 671-684.

Cronbach, L.J. (1975) Beyond the two disciplines of scientific psychology. American Psychologist, 30, 116-127.

Cronbach, L.J. (1976) On the design of educational measures. In D.N.M. de Gruijter, L.J. Th. van der Kemp & H.F. Crombag (Eds) Advances in psychological and education measurement. London: Wiley, pp. 199-208.

Cronbach, L.J. (1976b) Research on classrooms and schools. Formulations of questions, design, and analysis. Stanford, California: Occasional papers of the Stanford Evaluation Consortium.

Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972) The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley.

Cronbach, L.J., & Snow, R.E. (1977) Aptitudes and instructional methods. A handbook for research on interactions. New York: Irvington.

Cronbach, L.J., & Webb, N. (1975) Between-class and within-class effects in a reported aptitude x treatment interaction: Reanalysis of a study by G.L. Anderson. Journal of Educational Psychology, 67, 717-724.

Eysenck, H.J., & Eysenck, S.B.G. (1969) Personality structure and measurement. London: Routledge & Kegan Paul.

Gagné, R.M., & Cropper, G.L. (1965) Individual differences in learning from visual and verbal presentations. American Institutes for Research, Pittsburgh.

Guilford, J.P. (1967) The nature of human intelligence. New York: McGraw-Hill.

Gustafsson, J.-E. (1975) Interactions and the experimental research on teaching methods. Scandinavian Journal of Educational Research, 19, 45-57.

Gustafsson. J.-E. (1976) Verbal and figural aptitudes in relation to instructional methods. Studies in aptitude-treatment interactions. Göteborg: Acta Universitatis Gothoburgensis.

Gustafsson, J.-E. (1977) Differential effects of imagery instructions on pupils with different abilities. Scandinavian Journal of Educational Research, 21, 151-179.

Gustafsson, J.-E. (1978) A note on class effects in aptitude x treatment interactions, Journal of Educational Psychology, 70, 142-146.

Gustafsson, J.-E. (1979a) Suppression of visualization by reading as a function of aptitudes. Unpublished manuscript.

Gustafsson, J.-E. (1979b) Attitudes towards the school, the teacher and the classmates at the class level and the individual level. British Journal of Educational Psychology, in press.

Gustafsson, J.-E. (1979c) Testing and obtaining fit of data to the Rasch model. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 8-12 April.

Gustafsson, J.-E. (1979d) Pupil achievement and attitude as a function of teacher and pupil personality. Unpublished manuscript.

Gustafsson, J.-E., & Lindström, B. (1978) Describing and testing aptitude-treatment interaction effects with structural equation models: Reanalysis of a study by M.J. Behr. Reports from the Institute of Education, University of Göteborg, no. 73.

Härnqvist, K. (1968) Relative change in intelligence from 13 to 18. I. Background and methodology. Scandinavian Journal of Educational Research, 9, 50-82.

Jöreskog, K.G. (1969) A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 34, 183-202.

Jöreskog, K.G. (1970) A general method for analysis of covariance structures. Biometrika, 57, 239-251.

Jöreskog, K.G. (1971) Simultaneous factor analysis in several populations. Psychometrika, 36, 409-426.

Jöreskog, K.G. (1973) A general method for estimating a linear structural equation system. In A.S. Goldberger & O.D. Duncan (Eds.): Structural equation models in the social sciences. New York: Seminar Press, 85-112.

Jöreskog, K.G. (1974) Analyzing psychological data by structural analysis of covariance matrices. In R.C. Atkinson, D.H. Krantz, & R.D. Suppes (Eds.): Contemporary developments in mathematical psychology, Volume II. San Francisco: W.H. Freeman & Col., pp. 1-56.

Jöreskog, K.G. (1977) Structural equation models in the social sciences: Specification, estimation and testing. In P.R. Krishnaiah (Ed.): Application of statistics. North-Holland Publishing Co., 1977, 265-287.

Jöreskog, K.G. & Sörbom, D. (1976) Statistical models and methods for test-retest situations. Im D.N.M. de Gruijter, L.J.Th. van der Kamp & H.F. Crombag (Eds.): Advances in psychological and educational measurement. London: 135-137.

Jöreskog, K.G. & Sörbom, D. (1977) Statistical models and methods for analysis of longitudianl data. In D.J. Aigner & A.S. Goldberger (Eds.): Latent variables in socioeconomic models. Amsterdam: North-Holland Publishing Co., 285-325.

Jöreskog, K.G., & Sörbom, D. (1978) LISREL IV. A general computer program for estimation of linear structural equation systems by maximum likelihood methods. University of Uppsala, Department of Statistics.

Keesling, J.W. (1976) Components of variance multilevel analysis. Paper presented at the Conference on Methodology for Aggregating Data in Education Research, Stanford University, October 23-24.

Keesling, J.W. (1978) A multivariate components of covariance model for hierarchical data at two levels. Paper presented at the annual meeting of the American Education Research Association, Toronto.

Schmidt, W.C. (1969) Covariance structure analysis of the multivariate random effects model. Unpublished doctoral dissertation, University of Chicago.

Snow, R.E. (1978) Aptitude processes. In R.E. Snow, P.A. Federico W. Montague (Eds): Aptitude, learning and Instruction: Cognitive process analysis, in press.

Sörbom, D., (1974) A general method for studying differences in factor means and factor structures between groups. British Journal of Mathematical and Statistical Psychology, 27, pp. 229-239.

Sörbom, D. (1975) Detection of correlated errors in longitudinal data. British Journal of Mathematical and Statistical Psychology, 28, pp. 138-151.

Sörbom, D. (1976) A statistical model for the measurement of change in true scores. In D.N.M. de Gruijter and L.J.Th. van der Kamp (Eds.): Advances in psychological and educational measurement. New York: Wiley, pp. 159-169.

Sörbom, D. (1978) An alternative to the methodology for analysis of covariance. Psychometrika, 43, pp. 381-396.

Sörbom. D. (1979) Personal communication.

Werts, C.E., Jöreskog, K.G., & Linn, R.L. (1973) Identification and estimation in path analysis with unmeasured variables. Educational and Psychological Measurement, 78, 1469-1484.

Wiley, D.E. (1973) The identification problems for structural equation models with unmeasured variables. In A.S. Goldberger & O.E. Duncan (Eds): Structural equation models in the social sciences. New York: Seminar Press, pp. 69-83.

Winer, B,J, (1978) Statistics and data analysis: Trading bias for reduced mean squared error. In M.R. Rosenzweig, & L.W. Porter. (Eds): Annual Review of Psychology, Volume 29. Palo Alto: Annual Reviews Inc.